Interplay between HIV-1 Transcription and the Pathways of Intracellular Innate Defense

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Claudia Althaus

von Bretzwil, BL

Promotionskomitee

Prof. Dr. Alexandra Trkola (Vorsitz)

Prof. Dr. Marek Fischer (Leitung der Dissertation)

PD Dr. Karin Metzner (Leitung der Dissertation)

Prof. Dr. Huldrych Günthard

Prof. Dr. John Robinson

Zürich, 2011

Table of contents

Summary

Human immunodeficiency virus type 1 (HIV-1) infection causes a life-long chronic disease despite the tremendous success of combination antiretroviral therapy (cART). This is due to the ability of HIV-1 to persist in latent forms in long-lived infected cells. Currently approved antiretroviral drugs target multiple steps in the viral life cycle. Thus, latent cellular reservoirs of HIV-1 remain untouched and represent one of the major obstacles towards the goal of curing HIV-1 infection.

The mechanisms of induction and maintenance of HIV-1 latency have not been fully resolved so far. Besides regulation by transcription factors and/or epigenetic modulation of the proviral DNA, one hypothesis suggests that HIV-1 latency may be mediated by the intracellular machinery governing innate antiviral defences, such as RNA-interference. This mechanism is mediated by small noncoding RNAs, which bind to complementary mRNA and finally lead to cleavage of the mRNA or suppression of translation. HIV-1 latency could be induced by targeting viral or cellular RNA necessary for viral replication.

Chapter 1 describes the establishment of a method to measure viral RNAs with single copy sensitivity. Quantitative PCR (qPCR) using fluorescent hydrolysis probes (FH-probes) of variable HIV-1 genomes can result in underestimation of viral copy numbers due to mismatches in the FH-probe's target sequences. Particularly for detection of viral RNA in patients, successfully treated with cART, or in latently infected cells, highly sensitive qPCRs are essential. The study resulted in empirically validated novel principles of FH-probe design regarding conservation and qPCR performance. Several FH-probes used to quantify HIV-1 DNA over 6 orders of magnitude approached single copy sensitivity. Moreover, application of an algorithm to scan sequence databases for FH-probes with optimal phylogenetic conservation, allowed to identify functional FH-probes in various regions of the HIV-1 genome approaching coverage of the global HIV-1 pandemic.

The second chapter illustrates the development of a method to efficiently select and sequence small noncoding RNAs (sncRNAs), the key molecules of RNA interference. Since HIV-1 encoded sncRNAs represent only a small minority of sncRNAs in an infected cell, their detection remains difficult and reported frequencies of HIV-1

sncRNAs are usually <0.5%, if at all. Our approach was to enrich sequences homologous to HIV-1 by hybridization capture using HIV-1 ssDNA hybridization probes attached to Streptavidin beads. With this optimized selection method we were able to enrich low abundant HIV-1 sncRNAs more than 100-fold, i.e., >70% of captured sncRNAs were derived from HIV-1. This method was further applied to characterize the scope of viral sncRNA expression in primary HIV-1 infected cells. A multitude of HIV-1 sncRNAs was identified, distributed throughout the HIV-1 genome, yet, they clustered in contigs with hot-spots at distinct sites. Furthermore, hybrids of sense and antisense sncRNAs of one contig inhibit HIV-1 replication in primary macrophages demonstrating one possible function of these sncRNAs.

The third section refers to a study published in collaboration with colleagues from our research group. The study focused on monitoring HIV-1 RNA transcription patterns in peripheral blood mononuclear cells (PBMC) during cART in acutely HIV-1 infected patients to assess the effect of early treatment on cellular viral reservoirs (Schmid *et al*., 2010). I developed the appropriate qPCR protocols (please compare first chapter), assisted with the search for optimal patient matched primers and FH-probes for highly sensitive qPCRs, supervised the qPCR procedures, and helped in analyzing and interpreting the qPCR data. This study showed that early cART in acute HIV-1 infection significantly depleted the number of transcriptionally active proviruses when compared to levels detected in patients starting ART during chronic HIV-1 infection. Therefore, future studies aiming at HIV-1 eradication should be initiated in patients during acute infection.

In conclusion, my thesis mainly focused on the establishment of highly sensitive and improved methods for the detection of HIV-1 genomes and HIV-1 derived small noncoding RNAs, respectively, to enable detailed studies of various aspects of HIV-1 latency. Thus, quantitative PCR protocols were developed to detect a broad range of highly diverse HIV-1 variants with single copy sensitivity, and a novel approach for very efficient enrichment of low abundant HIV-1 sncRNAs was established. Both methods lay the foundation for further studies, not only in the field of HIV-1, but also in other research areas where highly sensitive methods for the detection of RNA and sncRNAs are needed.

Zusammenfassung

Eine Infektion durch das Humane Immunodefizienz-Virus vom Typ 1 (HIV-1) verursacht eine lebenslange chronische Erkrankung, trotz einer sehr erfolgreichen kombinierten antiretrovirale Therapie (*engl.* cART). Ein wesentlicher Grund ist die Fähigkeit von HIV-1 in eine Latenzphase einzutreten und sich somit für das Immunsystem unsichtbar zu machen. Die mit HIV-1 latent infizierten Zellen können Jahre, wenn nicht Jahrzehnte überleben, ehe sie in die produktive Phase übergehen. Die heute zugelassenen antiretroviralen Medikamente greifen an verschiedensten Stellen in den viralen Lebenzyklus ein. Das latente zelluläre Reservoir von HIV-1 bleibt jedoch bis zum heutigen Tage unangetastet und repräsentiert eines der grössten Hindernisse auf dem Weg zur Heilung einer HIV-1 Infektion.
Die Mechanismen der Induktion und Erhaltung der HIV-1 Latenz wurden bis heute noch nicht vollständig gelöst. Neben einer Regulierung durch Transkriptionsfaktoren und/oder epigenetische Modulationen der proviralen DNA, besagt eine dritte Hypothese, dass die HIV-1 Latenz durch intrazelluläre Mechanismen, welche für die angeborene antivirale Abwehr zuständig sind, verursacht werden könnte. Ein Beispiel für eine solche angeborene antivirale Abwehr ist die RNA Interferenz. Dieser Prozess wird durch kleine nicht-kodierende RNAs gesteuert, welche sich an komplementäre mRNAs binden und schlussendlich zu einer Zerteilung dieser mRNAs oder einer Unterdrückung der Translation führen. HIV-1 Latenz könnte daher durch eine gezielte Degradation von viraler oder zellulärer RNA, welche notwendig für die virale Replikation sind, verursacht werden.

Kapitel 1 beschreibt die Einführung einer Methode um virale RNAs mit der Sensitivität einer einzelnen Kopie messen zu können. Die quantitative PCR (qPCR) variabler HIV-1 Genome mit FH-Proben (*engl.* fluorescent hydrolysis probes) kann, im Falle einer Diskrepanz mit der Ziel-Sequenz der FH-Probe, zu einer Unterschätzung der viralen Kopienzahl führen. Besonders beim Nachweis viraler RNA in erfolgreich mit ART behandelter Patienten, oder in latent infizierten Zellen, sind höchst sensitive qPCRs entscheidend. Aus der Studie ergaben sich empirisch bestätigte neue Prinzipien - bezüglich Konservierung und qPCR Effizienz - zur Gestaltung von FH-Proben. Bei mehreren FH-Proben, welche zur Quantifizierung von HIV-1 DNA über 6 Grössenordungen angewendet wurden, lag die Sensitivität bei

einer Kopie. Zusätzlich ermöglichte der Gebrauch eines Algorithmus die Durchsuchung von Sequenz Datenbanken nach FH-Proben mit optimaler phylogenetischer Konservierung. Dies ermöglichte die Identifizierung von funktionellen FH-Proben an verschiedenen Stellen des HIV-1 Genoms, welche annähernd die globale HIV-1 Pandemie abdecken.

Das zweite Kapitel illustriert die Entwicklung einer effizienten Methode um kleine nicht-kodierende RNAs (*engl.* sncRNAs) zu selektionieren und sequenzieren, welche die Schlüsselmolekule der RNA Interferenz darstellen. Da HIV-1 kodierte sncRNAs nur eine kleine Minderheit der sncRNAs einer infizierten Zelle darstellen, bleibt der Nachweis schwierig, und die Häufigkeit dieser HIV-1 sncRNAs wird normalerweise als <0.5%, wenn überhaupt vorkommend, beschrieben. Unser Ansatz war HIV-1 homologe sncRNAs durch Hybridisierung mit HIV-1 DNA Hybridisierungs-Proben anzureichern. Mit dieser optimierten Methode war es uns möglich die selten vorkommenden HIV-1 sncRNAs mehr als 100-fach anzureichern, das heisst >70% der selektionierten sncRNAs stammten von HIV-1. Diese Methode wurde auch angewendet um die Bandbreite der viralen sncRNA Expression in primären HIV-1 infizierten Zellen zu charakterisieren. Eine Vielzahl von HIV-1 sncRNAs wurde identifiziert, welche auf dem ganzen HIV-1 Genom verteilt sind, und doch in gewissen Contigs in spezifischen Regionen gehäuft anzutreffen sind. Zusätzlich konnten wir zeigen, dass Hybride von sncRNAs mit positiver und negativer Polarität die HIV-1 Replikation in primären Makrophagen inhibieren, was eine mögliche Funktion dieser sncRNAs darstellen könnte.

Der dritte Teil bezieht sich auf eine Studie, die ich in Zusammenarbeit mit Kollegen von unserer Forschungsgruppe durchgeführt habe. Die Studie hatte die Aufzeichnung der HIV-1 RNA Transkriptions Muster in peripheren mononukleären Blutzellen (*engl.* PBMC) zum Fokus. Diese wurden während der medikamentösen Behandlung von akut HIV-1 infizierten Patienten gemessen, um den Effekt der frühen Behandlung auf das virale Reservoir abzuschätzen (Schmid *et al*., 2010). Meine Beträge waren das Erstellen der qPCR Protokolle (im ersten Kapitel beschrieben), die Assistenz bei der Suche nach optimal an die viralen Sequenzen der Patienten angepassten Primern und FH-Proben, sowie die Mithilfe bei der Analyse und Interpretation der Resultate der qPCRs. Die Studie konnte zeigen, dass frühe ART

während der akuten HIV-1 Infektion die Anzahl der transkriptionell aktiven Proviren signifikant senkt, verglichen mit Patienten, die die Behandlung erst während der chronischen Phase begannen. Daher sollten zukünftige Studie, welche die Eradikation von HIV-1 zum Ziel haben, bei Patienten während der akuten Phase initiiert werden.

Zusammenfassend war meine Arbeit fokussiert auf die Etablierung von sehr sensitiven und optimierten Methoden zum Nachweis von HIV-1 Genomen und kleinen nicht-kodierenden von HIV-1 stammenden RNAs, um detaillierte Studien zu verschiedenen Aspekten der HIV-1 Latenz durchführen zu können. Daher wurden quantitative PCR Protokolle entwickelt, um ein möglichst breites Spektrum stark unterschiedlicher HIV-1 Varianten mit grosser Sensitivität nachzuweisen. Auch wurde ein neuer Ansatz für eine sehr effiziente Anreicherung von selten vorkommenden HIV-1 sncRNAs gefunden. Beide Methoden legen die Grundlage für zukünftige Studien nicht nur im Feld von HIV-1, sondern auch in anderen Forschungsgebieten, in welchen hoch sensitive Methoden zum Nachweis von RNA und sncRNAs gebraucht werden.

Introduction

1. Human immunodeficiency virus type 1

1.1 History and Epidemiology

In the early 1980s, a mysterious disease pattern began to spread among men who have sex with men in New York[1], San Francisco and Los Angeles[2] causing low CD4 T-cell count, various bacterial infections and Kaposi's sarcoma (KS), an otherwise very rare skin cancer. A few years later, in 1983, Françoise Barré-Sinoussi and Luc Montagnier[3] reported to have found particles of a virus connected to this disease pattern named AIDS (acquired immunodeficiency syndrome). The virus is known today as human immunodeficiency virus type 1 (HIV-1).

Since the mid 1990s, effective treatment is available against HIV-1 by combining several antiretroviral drugs. Combination antiretroviral therapy (cART) is able to clear almost all extracellular viruses and significantly reduces morbidity and mortality of patients[4]. However, cART is not able to cure HIV-1 infection, as the virus hides in latently infected cells and immediately reactivates upon cessation of treatment.

Almost 30 years after the first mentioning of AIDS, in 2009, around 33 million people worldwide were infected with HIV-1; 2.6 million people became newly infected and 1.8 million people died of AIDS (UNAIDS, 2010). The HIV pandemic is caused by HIV-1. In 1986, HIV type 2 (HIV-2) was isolated from a West African patient with AIDS[5]. HIV-2 has a lower virulence and infectivity than HIV-1, and it is mainly prevalent in Western Africa[6].

1.2 Genome and Morphology

HIV-1 is a lentivirus and belongs to the family of retroviruses. Its genome is composed of two identical single-stranded RNA molecules (~9.2-9.4 kb) and nine genes. The proviral genome is flanked by the 5' and 3' LTRs (long terminal repeats) (Fig. 1a). The 5' LTR serves as a promoter for the RNA-polymerase II. HIV-1 encodes three structural and enzymatic genes, four accessory and two regulatory genes[7].

The HIV-1 RNA genome contains, like human cellular messenger RNA (mRNA), a 7-methylguanosine cap at the 5' end and is polyadenylated at the 3' end. The two LTRs are only present in parts in the HIV-1 RNA, namely the R (repeat) region is the only

repetitive sequence that exists at the 5' and the 3' end of the HIV-1 RNA (Fig. 1b). The R region forms a very strong and characteristic secondary structure, which is termed TAR (trans-activating response region) and serves as a binding site for the Tat (transactivator of transcription) protein[8,9]. The TAR element is one of the most structured regions, however, the whole HIV-1 RNA genome displays a extensive secondary structure[10]. Adjacent to the U5 region is the primer binding site (PBS), which is complementary to and binds the 3' end of host cellular $tRNA_{Lys}$[11,12]. The $tRNA_{Lys}$ functions as a primer for the reverse transcriptase to initiate the synthesis of the minus strand of the viral DNA. Around 700 nucleotides (nt) downstream of the start of transcription the dimerization initiation site (DIS) is located, which induces the dimerization of the 2 RNA copies that are incorporated into one virion. At the 3' end of the HIV-1 RNA, a polyadenylation signal is located in the R region and leads to synthesis of a poly(A)-tail, whereas this signal is suppressed at the 5' end[13].

The three structural and enzymatic genes are *gag* (group-specific antigen), *pol* (polymerase), and *env* (envelope) (Fig. 1b). *Gag* encodes four capsid proteins after cleavage: matrix (MA, p17), capsid (CA, p24), nucleocapsid (NC, p9), and p6 (Fig. 1c). The polyprotein Pol is processed into three enzymatic proteins – protease (PR), reverse transcriptase (RT), and integrase (IN). The RT transcribes viral single-stranded RNA (ssRNA) into DNA, which is integrated by integrase into the host genome and viral precursor poly-proteins are finally cleaved by protease into functional proteins[14]. *Env* (envelope) encodes gp160 which is cleaved to surface glycoprotein SU (gp120) and transmembrane glycoprotein TM (gp41) by the cellular protease furin[15]. The regulatory genes are *tat* (transactivator of transcription) and *rev* (regulator of virion expression). Tat acts as transcriptional activator through binding to the flanking region, the TAR region in the LTR[8,9]. Rev transports viral RNA from the nucleus to the cytoplasm[16-18]. The accessory genes of HIV-1 are *nef* (negative effector), *vif* (virion infectivity factor), *vpu* (viral protein u), and *vpr* (viral protein r). The Nef protein is important for the efficiency of virus replication[19]. Among other functions, Vif inhibits the function of APOBEC3G (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G), which is a deaminase with antiviral function[20].

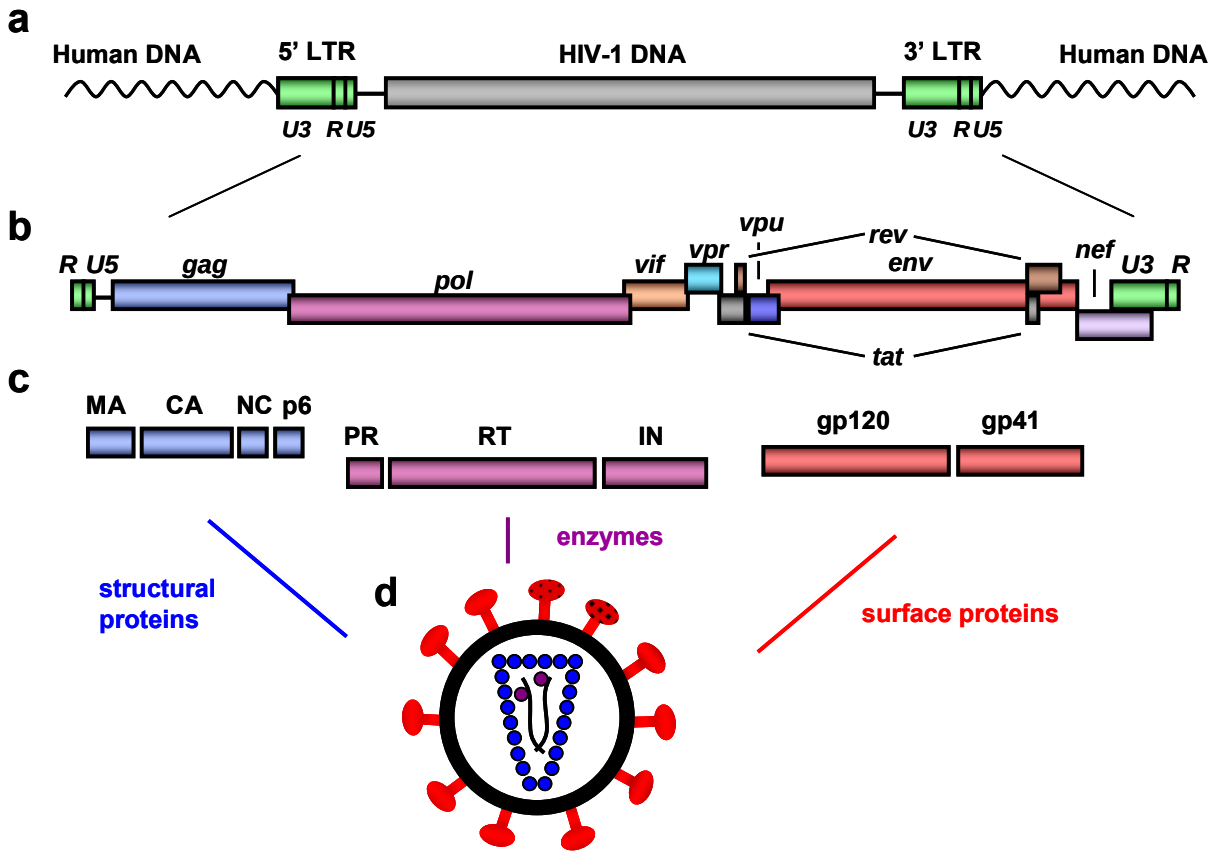

Figure 1: HIV-1 provirus, RNA genome, proteins and virion structure. a) The HIV-1 provirus is integrated into the human genome. Integrated HIV-1 DNA is flanked at the 5' and at the 3' end by 2 complete long terminal repeats (LTRs: U3-R-U5). b) Transcription of the HIV-1 RNA genome starts in the R region of the 5' LTR and ends with the R of the 3' LTR. HIV-1 encodes three structural and enzymatic (*gag, pol, env*), four accessory (*vif, vpu, vpr, nef*), and two regulatory genes (*tat, rev*). c) 3 genes of the HIV-1 genome encode structural proteins and enzymes: the Gag precursor poly-protein is processed into MA (matrix), CA (capsid), NC (nucleocapsid) and p6, *pol* encodes the enzymes PR (protease), RT (reverse transcriptase) and IN (integrase), while *env* is encoding the two surface glycoproteins, gp120 and gp41. d) The HIV-1 virion is surrounded by a lipid membrane derived from the host cell (black) in which HIV-1 Env proteins (red) are embedded. The 2 HIV-1 RNA copies are surrounded by nucleocapsid proteins. In addition, proteins necessary for reverse transcription and integration of the provirus are incorporated into the virion.

The structural proteins form the virion, which is about 100-130 nm in diameter (Fig. 1d). Its outer layer consists of a lipid membrane – derived from the host cell – in which the viral envelope proteins, gp120 and gp41, are embedded. These envelope proteins are essential for attachment and entry of the virion into the target cell via its receptor, CD4. The inner surface of the membrane is lined by the matrix. The capsid forms a conical core encircling the RNA, which is coated by nucleocapsid. In addition to viral proteins, the membrane and the interior of the virus contain proteins derived from the host cell[7,21].

1.3 Life cycle of HIV-1

When a HIV-1 virion infects a cell, homotrimeric Env proteins first bind to the cellular CD4 receptor and subsequently to a chemokine co-receptor, predominantly either CCR5 or CXCR4[22] (Fig. 2). This interaction results in conformational changes leading to fusion of the viral and the host cell membranes. The viral core enters the cell. The single-stranded RNA is converted to double-stranded DNA by the HIV-1 encoded reverse transcriptase. The DNA enters the nucleus and is integrated into the host genome by the viral integrase, forming the provirus[23].

In this stage, the provirus can hide in the infected cell for years without being transcribed or only being transcribed at low levels, which is then called the latent stage. During productive infection, the proviral DNA is transcribed by the cellular RNA polymerase II. The first transcripts that appear are mRNAs which encode the regulatory proteins Tat, Rev and Nef. Those mRNAs are multiply spliced by the cellular splicing machinery and, thus, are able to be exported into the cytoplasm[24]. The Rev and the Tat proteins subsequently return to the nucleus. Rev is responsible for efficient export of unspliced and singly spliced viral mRNAs to the cytoplasm[25]. Tat mediates transcriptional activation by binding to the TAR stem-bulge-loop[26]. Subsequently, in the late phase, unspliced and singly spliced mRNAs are produced and exported to the cytoplasm. Singly spliced mRNAs encode for the Env proteins and the accessory proteins Vif, Vpu and Vpr. From the unspliced RNA Gag and Pol proteins are translated[14]. Finally, 2 unspliced RNA molecules and several proteins assemble, and new virions are budding from the cell[27].

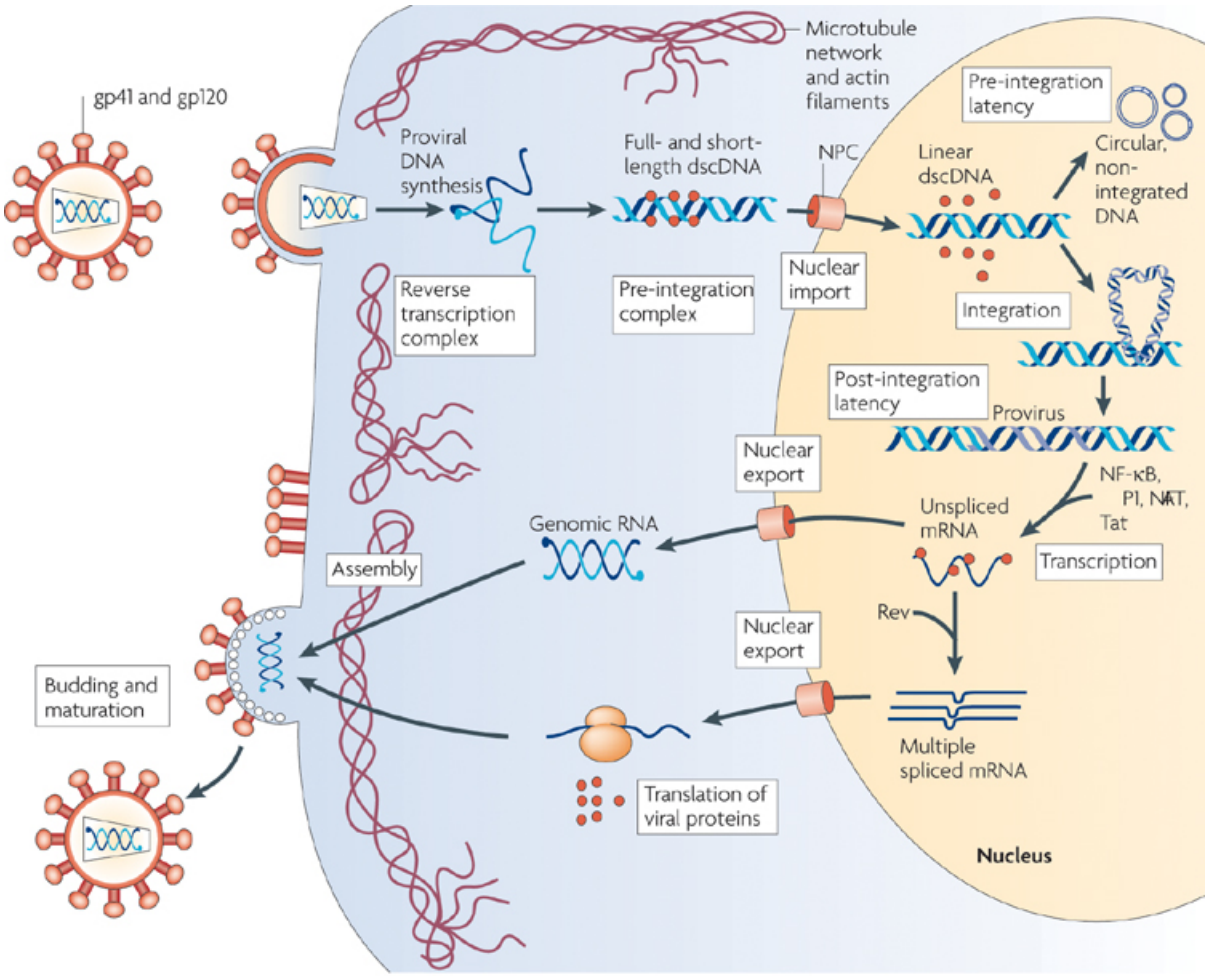

Figure 2: HIV-1 lifecycle. Viral fusion and entry requires the binding of glycoprotein gp120 to CD4 receptors at the cell surface as well as to chemokine co-receptors, either CCR5 or CXCR4. The viral core enters the cytoplasm and viral RNA is converted into proviral double-stranded DNA (dsDNA) by the reverse transcriptase. The pre-integration complex is transported into the nucleus through the nuclear pore complex (NPC), and the dsDNA either circularizes as one or two long terminal repeat-containing circles or is integrated into a host genome. After integration, the provirus can remain quiescent, existing in a permanent post-integration latent state. Upon activation, the viral genome is transcribed by the cellular RNA polymerase II. This process is influenced by cellular transcription factors (NF-κB, NFAT and SP1) and the viral transactivator Tat. The viral protein Rev regulates the export of unspliced or singly spliced viral mRNAs to the cytoplasm, which are translated into regulatory and structural viral proteins. New virions assemble and bud through the cell membrane, maturing through the activity of the viral protease. (adapted from Coiras et al., 2009 [28])

1.4 Transmission and course of infection

HIV-1 is transmitted through body fluids and an infection takes place through contact with contaminated blood or oral, anal, and vaginal sexual intercourse. HIV-1 positive women can infect their babies during pregnancy, birth, or through breast feeding[29,30].

The first phase of infection, the acute phase, normally lasts 4 to 8 weeks and is characterized by very high blood virus levels (viremia) and a dramatic decrease in numbers of $CD4^+$ T-cells[31]. The virus preferentially infects $CD4^+$ T-lymphocytes, as well as dendritic cells and macrophages. The adaptive immune system responds with formation of helper $CD4^+$ T-cells, $CD8^+$ cytotoxic T-cells and antibodies, which may partially clear the virus from the circulation, but do not wipe out the virus[32]. Nevertheless, this mounting of an HIV-1 specific immune response in the acute phase reduces the number of viral particles in the blood stream.

In the following asymptomatic phase, which lasts between 2 and 15 years, the virus constantly replicates and the $CD4^+$ T-cells decrease over time. The symptomatic phase, or AIDS stage, occurs, when the $CD4^+$ T-cell levels are low and the HIV-1 amount in the peripheral blood rises. At this stage, the immune system is immunodeficient and highly affective for opportunistic infections[30].

1.5 HIV-1 latency

Upon initiation of combined antiretroviral therapy (cART), virus replication is suppressed and the number of productively HIV-1 infected cells rapidly decrease[33,34]. But latently HIV-1 infected cells can survive for years if not decades during effective cART[35-37].

Latency is a powerful survival strategy that many viral and microbial pathogens have evolved to persist in the face of host defense[38]. By temporarily reducing gene expression at the cost of forsaking production of progeny, the pathogen can avoid destruction by the adaptive immunity, which recognizes mainly infected cells actively producing foreign antigens. In order to use latency as a valuable survival strategy, the pathogen will eventually have to reverse the condition of latency to a state of offspring production. Thus HIV-1 latency is reversible and in equilibrium with replication[36,37,39,40].

1.6 Suggested mechanisms inducing HIV-1 latency

A first model that attempts to explain the phenomenon of HIV-1 latency is that epigenetic modifications, such as chromatin remodelling, lead to a suppression of transcription (Fig. 3a and c). Tat is capable of remodelling chromatin in the LTR region to enhance the transcription rate[41-43]. One example of a host silencing mechanism efficiently downregulating HIV-1 expression is the recruitment of HDAC1

(histone deacetylase 1) to the HIV-1 LTR[42,44]. HDACs function to remove the acetyl group from the N-terminal histone tail, which leads to a condensation of the DNA structure and with that a repression of transcription. Consistent with a major role for HDACs in establishing HIV-1 latency, many drugs that inhibit HDAC activity, such as trichostatin A (TSA) and valproic acid (VPA)[45], are effective inducers of HIV-1 transcription in latently infected cells. Tat also counteracts this mechanism by recruiting HATs (histone acetyltransferases) to the LTR which acetylate the histone tails, and with that circumvent the transcriptional repression[46,47].

A second hypothesis is that HIV-1 latency is regulated by transcription factors (Fig. 3b and d). NF-κB is a cellular transcription factor, which is important for HIV-1 replication. The LTR harbours two binding sites for NF-κB. Upon activation, NF-κB is released by its inhibitor IκB and enters the nucleus, where it binds the LTR of HIV-1[48]. The NF-AT and PKC pathways also affect viral promoter activation in the HIV-1 LTR[49-51]. The sequestration of those cellular transcription factors in the cytoplasm of resting $CD4^+$ T-cells leads to a transcriptional silencing of HIV-1[52-54]. The most notable viral regulator of HIV-1 transcription is Tat, the trans-activator protein. Tat mediates transcriptional activation by binding to the TAR stem-bulge-loop[26]. There it mediates recruiting of essential transcription factors, such as the positive transcription-elongation factor b (P-TEFb)[55,56]. When Tat is absent or only present at low levels, short prematurely terminated non-polyadenylated transcripts are formed, which are not transferred to the cytoplasm[40]. In the presence of Tat full-length transcripts increase up to 100 fold[8,57,58]. However, in resting $CD4^+$ T-cells expression of Tat protein is inefficient and therefore Tat transactivation is presumably absent.

And a third model suggests that post-transcriptional gene silencing via RNA-interference is involved (Fig. 3e). Here, a protein complex is guided by a small RNA to the complementary mRNA. This interaction finally leads to degradation of the mRNA or translational repression.

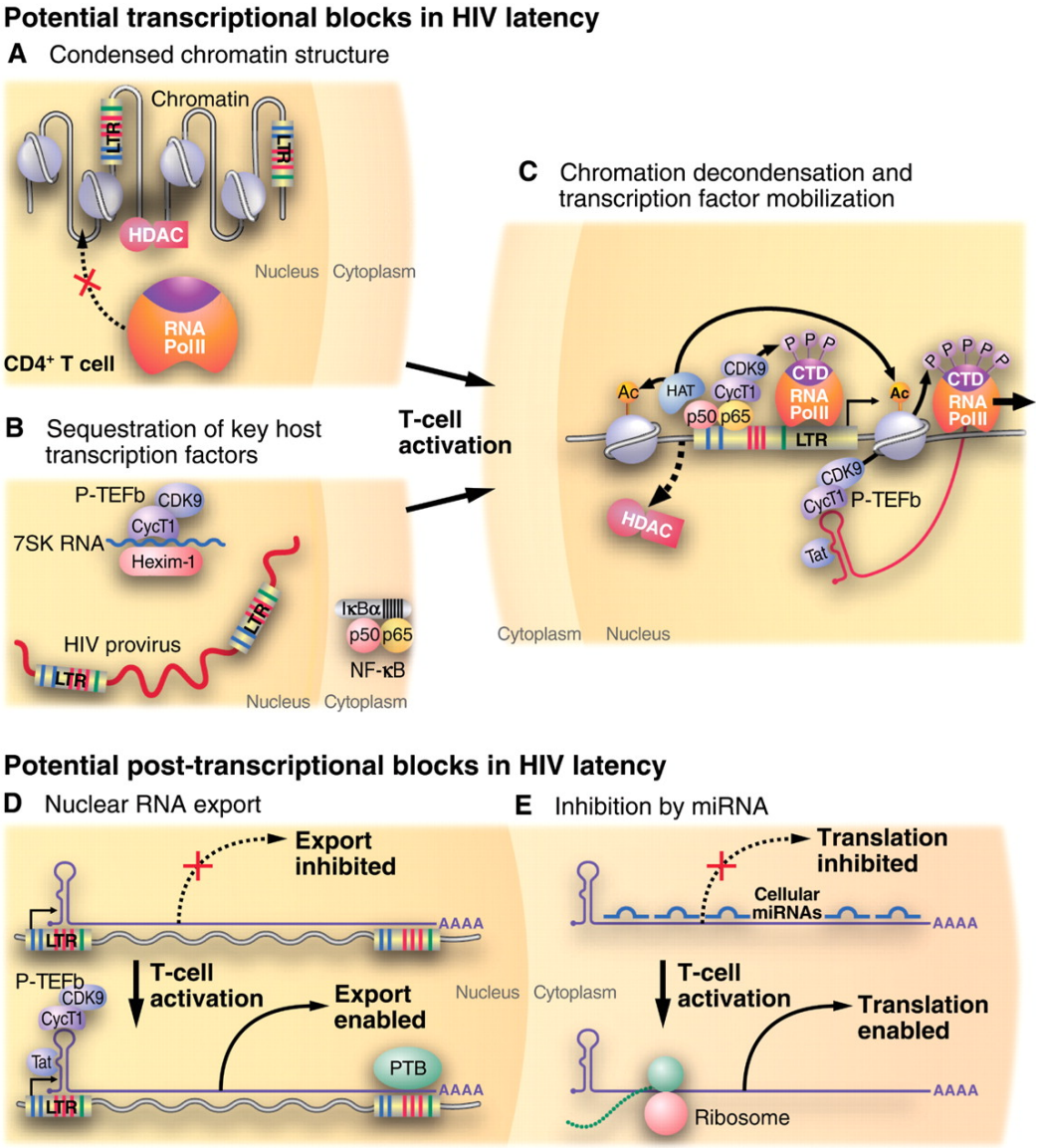

Figure 3. HIV-1 latency is the result of multiple restrictions on HIV-1 expression. a) Proviral latency is maintained, in part, by the action of several transcription factors that recruit, for instance, histone deacetylases (HDACs) to the HIV-1 LTR promoter, which results in condensation of chromatin and subsequently limits the ability of RNA polymerase to initiate transcription. b) Key cellular factors that are required for HIV-1 transcription, such as NF-κB or P-TEFb, are sequestered in resting $CD4^+$ T cells by cellular regulatory complexes. Release and mobilization of these factors is required for proviral expression. c) When histone acetyltransferases (HATs) overcome the effect of HDACs, coactivators such as NF-κB can recruit RNA polymerase complexes. Production of Tat allows the recruitment of P-TEFb, mediating an explosive increase in transcription and the escape of provirus from latency. d) The initial wave of Tat production may be restricted by inefficient export of multiply spliced HIV-1 mRNAs. Upon cellular activation those mRNAs are relieved by enhanced expression of polypyrimidine tract–binding protein (PTB). e) Cellular or possibly viral miRNAs that bind HIV-1 mRNAs may also restrict translation of early expressed HIV-1 mRNAs and so reduce Tat production. (adapted from Richman *et al.*, 2009 [59])

2. RNA-interference

The discovery of RNA interference was preceded by observations of transcriptional inhibition by antisense RNA expressed in transgenic plants[60]. More than 10 years later, Fire, Mello, and co-workers described a mechanism that is induced by small double-stranded RNAs and finally leads to suppression of protein expression[61]. Today, this process is called RNA interference (RNAi) and known to be mediated by regulatory small noncoding RNA (sncRNA).

2.1 Regulatory small noncoding RNAs

So far, there are three main classes of regulatory sncRNAs known to induce RNA interference: the microRNAs (miRNAs), the small interfering RNAs (siRNAs), and the PIWI-interacting RNAs (piRNAs). These sncRNAs share the characteristics of having a short size (~20-30 nt) and being associated with Argonaute (Ago)-family proteins[62]. The first miRNA, *lin-4*, was identified in *Caenorhabditis elegans*[63]. It plays a key role in regulating postembryonic developmental events[63]. Afterwards, miRNAs were also found in plants and other animals, including mammals. They were shown to influence a variety of biological processes by RNA-interference, for instance, embryogenesis, apoptosis, cell proliferation, and carcinogenesis[64]. Today, over 1,000 miRNAs (www.mirbase.org) are reported alone in human cells, which play a crucial role in the regulation of genetic and epigenetic processes. The second group are formed by siRNAs that are derived from long double stranded RNA and bind to the target mRNA with perfect sequence similarity. They can be either of endogenous, i.e., base-paired sense and antisense transcripts, or of exogenous, mainly viral, origin[62]. The third group of regulatory sncRNAs are the piRNAs[65]. They are derived from long single-stranded RNAs and are mainly expressed in germ cells, however, their existence could recently be shown also in somatic cells[66].

2.2 Biogenesis and function of regulatory small noncoding RNAs

MicroRNAs are encoded in the genome and their precursors are transcribed by the RNA polymerase II, resulting in long either completely or partially double-stranded primary transcripts[67]. Those are trimmed by microprocessors, protein complexes consisting of Drosha, an endonuclease with dsRNA specific activity, and the RNA binding protein DGCR8[68,69]. The precursor miRNAs (pre-miRNAs) are actively

exported to the cytosol by Exportin-5[70]. There they are processed by Dicer, another double strand specific endonuclease, into duplexes of 18-25 basepairs (bp)[71]. Those duplexes then bind the RNA induced silencing complex (RISC) whereby one strand is integrated into the protein complex and the other is been degraded[72]. This complex of the mature miRNA bound to RISC is then responsible for RNA-interference, by either cleaving the complementary target mRNA or repressing the translation.

SiRNAs are produced in a similar way. They are cleaved by the same endonuclease Dicer, and also induce RNA-interference by being incorporated into RISC. However, in contrast to the miRNAs, they are derived from complementary double stranded RNA progenitors, and ultimately rather lead to degradation of mRNA than to translational repression[73].

MiRNAs and siRNAs are not only able to lead to posttranscriptional degradation of mRNA or repression of translation but may also induce transcriptional gene silencing by epigenetic modifications[74]. Transcriptional gene silencing acts at the level of DNA interactions with promoters and can result in long-term silencing. The most common epigenetic modifications induced by sncRNAs are DNA methylation and histone modifications[74].

The third group of sncRNAs, piRNAs, are 24-31 nt long and silence transposons in animal germ cells. In contrast to siRNAs and miRNAs, they neither require Drosha nor Dicer for their biogenesis. Thus, piRNAs are single-stranded RNAs without going through a double-stranded intermediate state[75,76].

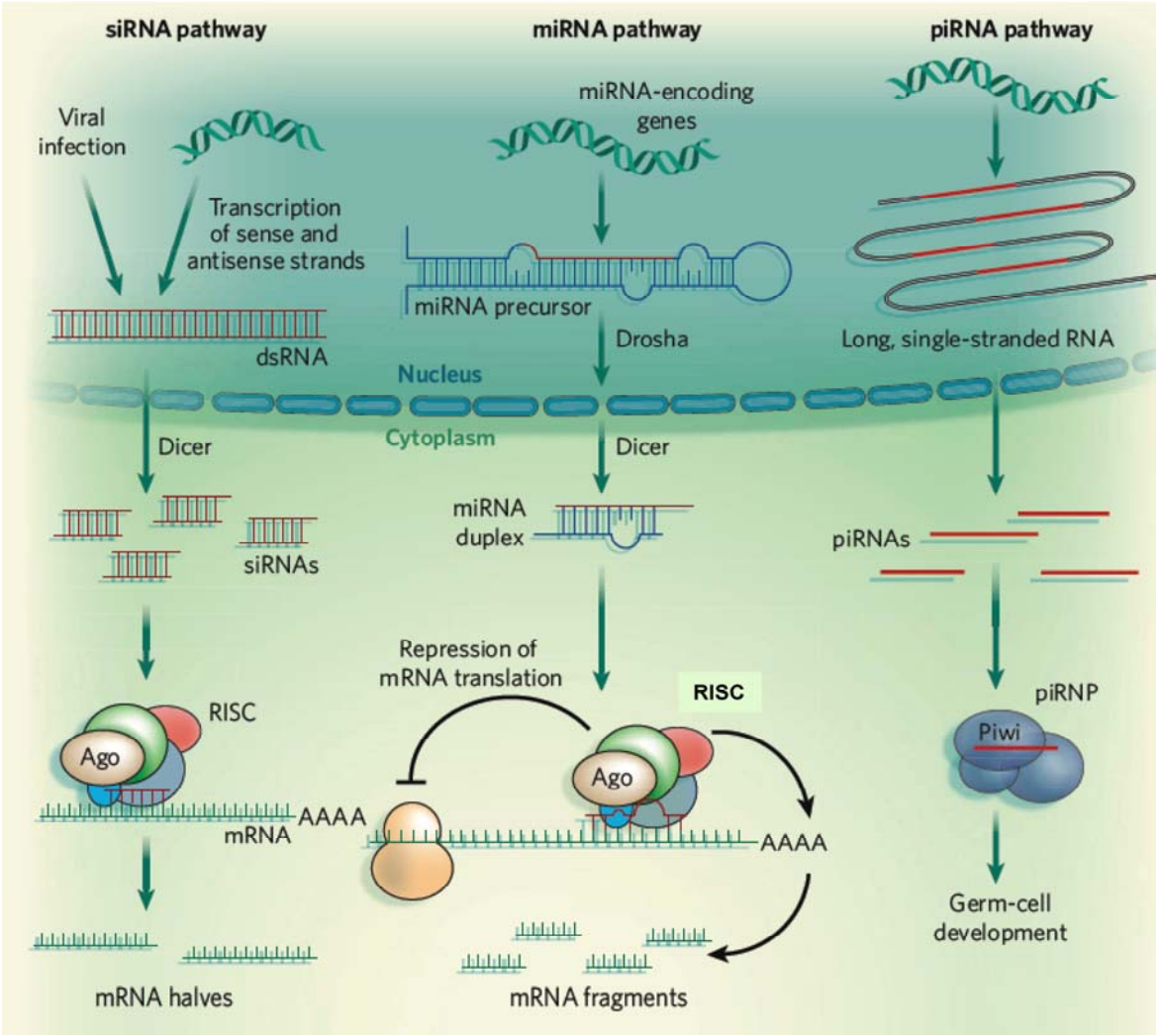

Figure 4: Formation and function of small noncoding RNAs. a) SiRNAs are processed from complementary double-stranded RNAs (dsRNAs). The enzyme Dicer cleaves dsRNA into shorter double-stranded siRNAs of about 20 bp. One strand then assembles into the RNA-induced silencing complex (RISC). This complex uses the siRNA guide to identify mRNAs with a perfect sequence complementary to the siRNA, leading to cleavage of the mRNA. b) MiRNAs are processed from genome-encoded precursors, which fold into intramolecular hairpins containing imperfectly base-paired segments. The processing generally occurs in two steps, and is catalyzed by the enzymes Drosha (in the nucleus) and Dicer (in the cytoplasm). One strand of the resulting miRNA duplex, resembling a siRNA, then incorporates into the RISC-complex. Depending on the level of sequence similarity, miRNAs induce mRNA degradation or repress their translation. c) PiRNAs are generated from long, single-stranded precursors in a process independent of Drosha and Dicer. These small RNAs associate with a subfamily of Argonaute proteins called Piwi proteins. This complex is thought to silence transposons in animal germ cells. (adapted from Großhans and Filipowicz, 2008 [77]).

2.3 Viral microRNAs

In 2004, the first viral miRNAs were described, which were discovered in Epstein-Barr virus (EBV) infected human B-cells[78]. To date, more than 140 viral miRNAs (vmiRNAs) have been discovered in herpesviruses[79,80] alone, but vmiRNAs were also identified in polyomaviruses (SV-40)[81] and retroviruses (HIV-1)[82-85].

Of interest, Kaposi's sarcoma-associated herpesvirus (KSHV) encodes as much as 12 pre-miRNAs that result in 17 distinct mature vmiRNAs. Those vmiRNAs may play critical roles in establishment and/or maintenance of KSHV latent infection[86]. EBV expresses vmiRNAs that target the viral protein LMP1 (latent membrane protein 1), which is a transmembrane protein essential for EBV-mediated growth transformation[87]. They were also described to be characteristic for type III latency in infected B lymphocytes[88]. In general, viral miRNAs in herpesviruses appear to be involved in latency/lytic switch, immune evasion, cell survival and proliferation[89].

3. Relationship between HIV-1 and RNA-interference

3.1 Cellular microRNAs interact with HIV-1

In 2005, Yeung and colleagues detected changes in cellular miRNA expression profiles, i.e., downregulation of a large pool of miRNAs in human HeLa cells transfected with HIV-1 plasmid pNL4-3[90]. A second study explored in more detail the importance of the miRNA pathway in the control of HIV-1 replication[91]. Using siRNAs against Drosha and Dicer in peripheral blood mononuclear cells (PBMC) from HIV-1 infected patients, they noticed faster virus replication kinetics in Drosha- or Dicer-depleted cells, as compared to cells treated with control siRNAs. This result was confirmed in latently infected U1 cells; downregulation of Drosha and Dicer led to an enhancement of virus replication. Therefore, both Drosha and Dicer contribute to the suppression of HIV-1 replication, and potentially are important factors for induction and maintenance of HIV-1 latency.

In the same study, Triboulet and colleagues also demonstrated that HIV-1 infection was associated with either up- or down-regulation of cellular miRNA clusters[91]. As an example, the miR-17/92 cluster encoding 7 miRNAs was significantly decreased in HIV-1 infected cells. Some miRNAs of this cluster, miR-17-5p and miR-20, may target the histone acetyltransferase (HAT) and the HIV-1 Tat cofactor PCAF (p300/CBP-asscociated factor), thus, regulate gene expression[91]. Several other studies explored also possible roles of cellular miRNAs to influence HIV-1 replication and demonstrated that various cellular miRNAs interfere with HIV-1 gene expression and inhibit viral replication[92-94]. Especially of interest is the study of Huang and co-workers[94] showing that the 3' end of the HIV-1 mRNAs is targeted and repressed by a cluster of cellular miRNAs including miR-28, miR-125b, miR-150, miR-223, and miR-382, and that those are enriched in resting $CD4^+$ T-cells as compared to activated $CD4^+$ T-cells. Specific inhibitors of these miRNAs substantially counteracted their effects on the target mRNAs. Those results indicate that cellular miRNAs may play a crucial role in HIV-1 latency.

In addition, cellular miRNA expression levels were shown to correlate with susceptibility of monocytes and macrophages to HIV-1 infection[95]. Freshly isolated monocytes from peripheral blood express significantly higher levels of miRNAs (miR-28, miR-150, miR-223, and miR-382) that possibly counteract HIV-1 than monocyte-derived macrophages. Suppression of these anti-HIV-1 miRNAs in monocytes was

shown to facilitate HIV-1 infectivity, whereas increase of the anti-HIV-1 miRNA expression in macrophages reduced HIV-1 replication.

3.2 HIV-1 counteracts cellular RNA-interference

The fact that HIV-1 transcription is active and produces new particles in the presence of RNAi components suggests that HIV-1 is able to counteract this inhibiting mechanism. Various viruses are known to express suppressor of RNA-silencing proteins, including NS1 (non-structural protein 1) of influenza[96] and Tas (transcriptional transactivator) of primate foamy virus[97]. The HIV-1 transactivator Tat was shown to antagonize the intracellular defence system by binding and inhibiting its key factor Dicer, which results in shutdown of the RNAi pathway[82]. Thus, Tat expression may prevent intracellular innate responses against virus infection, which could otherwise inhibit expression of viral mRNAs or proteins. These findings could not be confirmed when Tat was used in physiologically relevant levels[98,99].

Another report demonstrated that TRBP (TAR RNA binding protein), an important cofactor of Dicer in the RISC complex, interacts with Tat and its target sequence TAR, thus, being an essential component of the viral transactivation complex[91]. Hence, HIV-1 may hijack cellular proteins, as for instance TRBP, which are subsequently not available anymore for the RNA-interference pathway.

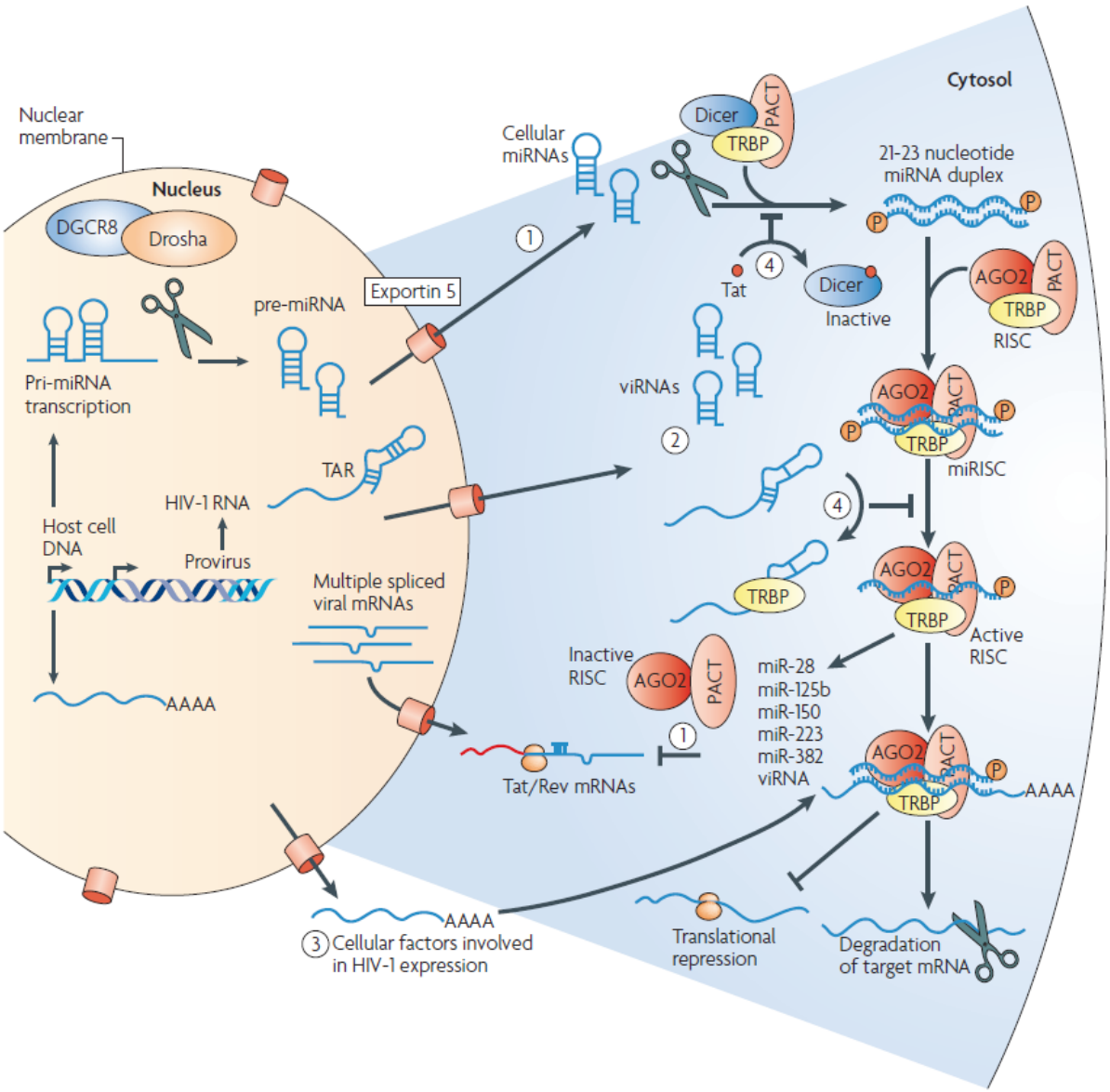

Figure 5. The microRNA pathway can modulate both cellular and viral gene expression. Potential mechanisms of the miRNA pathway to induce and maintain HIV-1 latency . 1) Cellular miRNAs can target and degrade viral mRNA or repress their translation. 2) The HIV-1 genome may produce miRNAs, called vmiRNAs, which can target viral RNAs and cellular mRNAs. 3) Cellular miRNAs can target cellular factors that are involved in HIV-1 replication. 4) Finally, viral products can interfere directly with RNA-interference machinery by hijacking Dicer (mediated by Tat) and TRBP (mediated by TAR). (adapted from Coiras *et al.*, 2009 [28])

3.3 HIV-1 encodes viral microRNAs

Using a computational method designed to find characteristic secondary structures in the nucleotide sequences, five candidate pre-miRNAs encoded by different regions of the HIV-1 genome were predicted[100]. So far, one of them has been experimentally confirmed in HIV-1 infected cells[85,101]. Furthermore, two HIV-1 encoded miRNAs and one siRNA have been discovered in HIV-1 infected cells[82,84,102]. Those HIV-1 sncRNAs are depicted in figure 6.

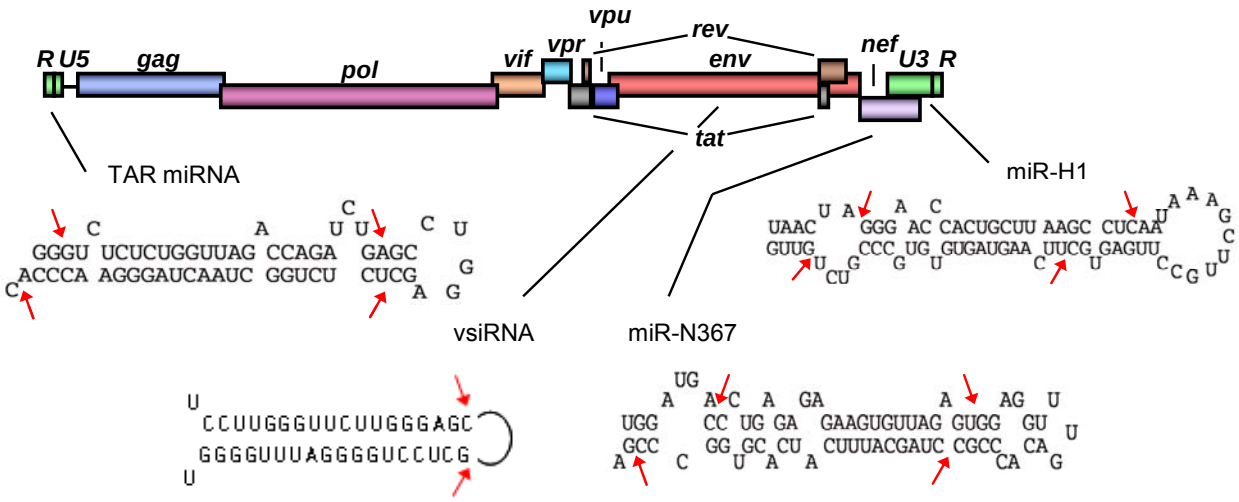

Figure 6. HIV-1 derived miRNAs and siRNAs. Positions, predicted secondary structures and proposed cleavage sites of Dicer (red arrows) of the TAR miRNAs[85,101], the viral siRNA (vsiRNA)[82], the miRNA derived from *nef* (miR-N367)[84], and the HIV-miR-H1[102] located in the U3, are displayed.

The first viral miRNA identified was detected in persistently HIV-1 infected MT-4 T cells and was derived from the *nef* region[84] (Fig. 6). This miRNA, called miR-N367, was detected by Northern blot analysis with probes directed to the *nef* region. The RNA fraction containing small RNA molecules (~25 nt) was subsequently cloned and sequenced, confirming the existence of miR-N367. Overexpression of miR-N367 did lead to suppression of HIV-1 LTR-driven transcription in reporter gene assays[83] suggesting that this miRNA has the ability to downregulate HIV-1 transcription. Additionally, this miRNA was able to downregulate nef expression in the subcapsular regions of the spleen in mice, as determined by immunofluorescence and RT-PCR[84]. One year later, Bennasser and co-workers identified a viral siRNA (vsiRNA) in HIV-1 infected cells[82] (Fig. 6). The viral siRNA derived from the Rev responsive element

(RRE) region within the *env* gene was detected with a specific probe by Northern blot analysis, showing a 21 nt long fragment not seen in uninfected cells. Consistent with its viral origin, the intensity of the 21 nt band increased over the course of HIV-1 infection in T-cells. Overexpression of this viral siRNA significantly reduced *env* mRNA levels and viral replication. Furthermore, co-expression of 2' O-methyl oligonucleotides complementary to the vsiRNA rescued *env* RNA expression.

The TAR element in the LTR with its characteristic hairpin structure was reported to be cleaved by Dicer to generate viral miRNAs, miR-TAR-5p and miR-TAR-3p[85,101] (Fig. 6). MiR-TAR-3p seemed to be more abundant in HIV-1 infected cells[103], and exhibited better suppression of HIV-1 gene expression[85]. However, both miRNAs might direct histone deacetylase HDAC-1 to the HIV-1 LTR promoter to silence transcription by chromatin remodelling[101]. Later, Klase and co-workers even proposed the omnipresent expression of TAR miRNA in all stages of infection, suggesting an anti-apoptotic function for infected cells[103]. Interestingly, a 124-mer TAR RNA domain present in HIV-2 is considered to possibly encode pre-miRNAs, and the resulting miRNAs may target a wide variety of cellular targets[104].

Another miRNA has been reported to be derived from a 81 nt structured region located immediately downstream of the 2 NF-kB sites in the LTR[102], called miR-H1 (Fig. 6). This miRNA supposedly has the ability to degrade the apoptosis antagonizing transcription factor (AATF) mRNA, resulting in downregulation of AATF gene expression and leading to cell death[102]. Furthermore, downregulation of AATF by miR-H1 was responsible for the suppression of genes coding for other proteins, including Dicer, as well as lower cell viability[105]. Additionally, mir-H1 seems to downregulate expression of cellular miRNA miR-149, which is considered to be able to target the HIV-1 *vpr* gene. Thus, the authors propose an epigenetic pathway responsible for the modulation of various cellular genes. In addition to those *in vitro* studies, miR-H1 variability in AIDS patients has been examined. A strong relationship between the miRNA sequence and the development of HIV-1-associated dementia (HAD) was detected[106].

Recently, various small noncoding RNAs with the length of miRNAs were discovered by ultra-deep sequencing, which are distributed over the whole HIV-1 genome[107]. A total of 47,773 clones were sequenced, of which 125 (0.3%) were HIV-1 specific sncRNAs. Especially a small 18 nt long RNA molecule, which is complementary to the HIV-1 primer binding-site (PBS), seemed to be highly abundant in productively

infected MT-4 T-cells. Supposedly, this small RNA is produced by Dicer cleavage of the HIV-1 PBS – $tRNA_{Lys}$ complex. Using immunoprecipitation, this small RNA was found to be associated with the Argonaute 2 (Ago2) protein, a central component of RISC. Together with the finding that this sncRNA led to a dose-dependent inhibition of virus production, this possibly suggests a function of this sncRNA in the RNA-interference pathway targeting HIV-1.

In contradiction to these reports, Pfeffer and colleagues detected only two viral sncRNAs in a total of 1,540 clones from HIV-$1_{HXB2-Bru}$ infected HeLa-T4+ cells[80]. However, no stable base-paired structures, that could display the potential precursor miRNAs, could be assigned to these HIV-1 specific sncRNAs and they were accordingly classified as degradation products by the authors[80]. Similarly, Lin and Cullen were not able confirm the existence of HIV-1 encoded miRNAs and siRNAs in T cell lines (MT-2, ACH-2) infected with HIV-1_{NL4-3} or the isolate QH0515 by Northern blot analysis and cDNA cloning[98]. However, the authors indicate that it may be possible for viral miRNAs to be present at very low levels (<0.5% of the total cellular miRNAs). It is very likely that expression of viral miRNAs vary between different cell types and different stages of infection.

1. Development of single copy sensitive qPCR for the detection of viral RNAs

Published as Althaus *et al*., J. Virol Methods. 2010 May;165(2):151-60

Journal of Virological Methods 165 (2010) 151–160

Contents lists available at ScienceDirect

Journal of Virological Methods

journal homepage: www.elsevier.com/locate/jviromet

Rational design of HIV-1 fluorescent hydrolysis probes considering phylogenetic variation and probe performance

Claudia F. Althaus[a,1], Sara Gianella[a,1], Philip Rieder[a,1], Viktor von Wyl[a], Roger D. Kouyos[b], Barbara Niederöst[a], Adrian Schmid[a], Karin J. Metzner[a], Beda Joos[a], Huldrych F. Günthard[a], Marek Fischer[a,*]

[a] Division of Infectious Diseases, University Hospital Zurich, University of Zurich, Switzerland
[b] Institute of Integrative Biology, ETH Zurich, Switzerland

Article history:
Received 17 September 2009
Received in revised form 13 January 2010
Accepted 20 January 2010
Available online 29 January 2010

Keywords:
HIV-1
Phylogeny
qPCR
Taqman probes

A B S T R A C T

Quantitative PCR (qPCR) using fluorescent hydrolysis probes (FH-probes; TaqMan®-probes) of variable genomes, such as HIV-1, can result in underestimation of viral copy numbers due to mismatches in the FH-probe's target sequences. Therefore both target conservation and physical properties of FH-probes, such as melting temperature, baseline fluorescence and secondary structure, should be considered in design of FH-probes.

Analysis of a database of 1242 near full-length HIV-1 sequences with a novel computational tool revealed that the probability of target and FH-probe identity decreases exponentially with FH-probe length. In addition, this algorithm allowed for identification of continuous sequence stretches of high conservation, from which FH-probes with global HIV-1 clade coverage could be chosen. To revise the prerequisites of physical FH-probe function, properties of 30 DNA and 21 chimeric DNA locked nucleic acid (DLNA) HIV-1 FH-probes were correlated with their performance in qPCR. This identified the presence of stable secondary structures within FH-probes and the base composition and thermal stability of the 5′ proximal end as novel predictors of FH-probe performance.

Thus, empirically validated novel principles of FH-probe design regarding conservation and qPCR-performance were identified, which complement and extend current rules for FH-probe design.

1. Introduction

Monitoring qPCR by FH-probes encounters its major pitfalls in assays examining phylogenetically diverse viruses such as HIV-1. While primers are relatively forgiving in respect to target variation (Christopherson et al., 1997), as little as one mismatch in a FH-probe's target sequence can greatly perturb the results of qPCR by underestimating copy numbers (Damond et al., 2007). This feature of FH-probes is widely exploited in genotyping of single nucleotide polymorphisms (Livak, 1999; Ranade et al., 2001).

Further factors influencing the function of FH-probes are related to their physical properties, which are determined by primary and secondary nucleic acid structure. Hybridization characteristics and kinetics of the FH-probe to its target have been assumed to be crucial for the performance of FH-probes. Hence, to ensure optimal binding of FH-probe to target, a choice of binding sites devoid of secondary structures and with 40–60% G/C content has been proposed (Bruijnesteijn Van Coppenraet et al., 2004; Gut et al., 1999; Livak et al., 1995; Mackay et al., 2002; Malnati et al., 2008). In addition, to ensure hybridization of the FH-probe before elongation of the newly synthesized DNA strand, it has been postulated that the melting temperature (T_m) of FH-probes should be 5–10 °C higher than the T_m of the primer binding to the same DNA strand as the FH-probe (Gut et al., 1999; Mackay et al., 2002).

Performance of FH-probes in qPCR assays also depends on intensity of fluorescence, which is suppressed in the absence of PCR target by a quenching moiety commonly placed at the 3′ end of the FH-probe and elicited by nucleolytic detachment of the base

Abbreviations: C_t, cycle threshold; Delta-T_m, T_m-difference of FH-probe and the primer binding the same strand; DLNA, chimeric DNA locked nucleid acid; FH-probe, fluorescent hydrolysis probe; qPCR, quantitative PCR; RT-qPCR, reverse transcriptase qPCR; SNR, signal to noise ratio; T_m, melting temperature.

* Corresponding author at: University of Zurich, University Hospital Zurich, Division of Infectious Diseases and Hospital Epidemiology, Rämistrasse 100, CH-8091 Zurich, Switzerland. Tel.: +41 44 255 36 10; fax: +41 44 255 32 91.
E-mail addresses: claudia.althaus@usz.ch (C.F. Althaus), sara.gianella@nadnet.ch (S. Gianella), philip.rieder@usz.ch (P. Rieder), viktor.vonwyl@usz.ch (V. von Wyl), roger.kouyos@env.ethz.ch (R.D. Kouyos), barbara.niederoest@usz.ch (B. Niederöst), adrian.schmid@usz.ch (A. Schmid), karin.metzner@usz.ch (K.J. Metzner), beda.joos@usz.ch (B. Joos), huldrych.guenthard@usz.ch (H.F. Günthard), marek.fischer@usz.ch (M. Fischer).
[1] These authors contributed equally to this work.

0166-0934/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jviromet.2010.01.012

Own contributions

My major contributions to this study were in both the practical and the analytical part. I carried out the measurement of the signal to noise ratios (SNRs) and the melting temperature (Tm) of all 51 FH-probes. I also did the standard curves with the selected FH-probes over a range from 3 million to 0.3 copies per PCR to determine the 50% endpoint of PCR-positive dilutions as an indicator of assay-sensitivity. Additionally, I generated the melting profile of an overlapping 23-mer DNA and a 13-mer DLNA (DNA locked nucleic acid) FH-probe, which confirmed that the DLNA FH-probe showed indeed a higher thermal stability. I analyzed the data, performed the statistical analyses for determination of predictors of FH-probe performance, and wrote the manuscript.

Rational design of HIV-1 fluorescent hydrolysis probes considering phylogenetic variation and probe performance

Claudia F. Althaus*[a], Sara Gianella*[a], Philip Rieder*[a], Viktor von Wyl[a], Roger D. Kouyos[b], Barbara Niederöst[a], Adrian Schmid[a], Karin J. Metzner[a], Beda Joos[a], Huldrych F. Günthard[a§] and Marek Fischer[a§]

* These authors contributed equally to this work, [a] Division of Infectious Diseases, University Hospital Zurich, University of Zurich, Switzerland, [b] Institute of Integrative Biology, ETH Zurich, Switzerland, § Corresponding authors

Quantitative PCR (qPCR) using fluorescent hydrolysis probes (FH-probes; TaqMan®-probes) of variable genomes, such as HIV-1, can result in underestimation of viral copy numbers due to mismatches in the FH-probe's target sequences. Therefore both target conservation and physical properties of FH-probes, such as melting temperature, baseline fluorescence and secondary structure, should be considered in design of FH-probes.

Analysis of a database of 1242 near full-length HIV-1 sequences with a novel computational tool revealed that the probability of target and FH-probe identity decreases exponentially with FH-probe length. In addition, this algorithm allowed for identification of continuous sequence stretches of high conservation, from which FH-probes with global HIV-1 clade coverage could be chosen. To revise the prerequisites of physical FH-probe function, properties of 30 DNA and 21 chimeric DNA locked nucleic acid (DLNA) HIV-1 FH-probes were correlated with their performance in qPCR. This identified the presence of stable secondary structures within FH-probes and the base composition and thermal stability of the 5'-proximal end as novel predictors of FH-probe performance.

Thus, empirically validated novel principles of FH-probe design regarding conservation and qPCR-performance were identified, which complement and extend current rules for FH-probe design.

Keywords

HIV-1, phylogeny, qPCR, taqman probes

Abbreviations

Ct, cycle threshold; Delta-Tm, Tm-difference of FH-probe and the primer binding the same strand; DLNA, chimeric DNA locked nucleid acid; FH-probe, fluorescent hydrolysis probe; qPCR, quantitative PCR; RT-qPCR, reverse transcriptase qPCR; SNR, signal to noise ratio; Tm, melting temperature

1. Introduction

Monitoring qPCR by FH-probes encounters its major pitfalls in assays examining phylogenetically diverse viruses such as HIV-1. While primers are relatively forgiving in respect to target variation[1], as little as one mismatch in a FH-probe's target sequence can greatly perturb the results of qPCR by underestimating copy numbers[2]. This feature of FH-probes is widely exploited in genotyping of single nucleotide polymorphisms[3,4].

Further factors influencing the function of FH-probes are related to their physical properties, which are determined by primary and secondary nucleic acid structure. Hybridization characteristics and kinetics of the FH-probe to its target have been assumed to be crucial for the performance of FH-probes. Hence, to ensure optimal binding of FH-probe to target, a choice of binding sites devoid of secondary structures and with 40-60% G/C content has been proposed[5-9]. In addition, to ensure hybridisation of the FH-probe before elongation of the newly synthesized DNA strand, it has been postulated that the melting temperature (Tm) of FH-probes should be 5-10°C higher than the Tm of the primer binding to the same DNA strand as the FH-probe[5,6].

Performance of FH-probes in qPCR assays also depends on intensity of fluorescence, which is suppressed in the absence of PCR target by a quenching moiety commonly placed at the 3' end of the FH-probe and elicited by nucleolytic detachment of the base carrying the fluorophore, usually placed at the FH-probe's 5' end. The physical proximity of fluorophore and quencher in the absence of PCR product has been reported to influence performance of FH-probes[9]. Optimal proximity of the two modified bases in the free FH-probe may be attained by keeping the FH-probe length minimal[10] or, if long FH-probes need to be designed, by attachment of the quencher to internal positions[11].

Another important rule advises to avoid guanosine at the residue carrying the fluorophore because it quenches fluorescence even after hydrolysis and it may be cleaved off with reduced efficiency[9,12].

In the present study, the phylogenetic constraints for design of FH-probes were addressed by analyzing the relationship of FH-probe match frequency with FH-probe length and single base conservation. Physical properties of FH-probes and their

influence on FH-probe performance were examined by testing 51 probes targeting 6 amplica (plural of amplicon) within the HIV-1 genome.

2. Methods

2.1. Conservation of FH-probes

The conservation of the chosen amplica was calculated by taking into account all circulating HIV-1 groups (subtypes M, N, O, CPZ). The sequences were obtained from curated alignments in the Los Alamos HIV Sequence Database (download 2007, www.hiv.lanl.gov) comprising a total of 1242 sequences including 17% subtype-B isolates. Due to location at the ends of the genome, the *early* and the *nef* amplica, respectively, were only covered by 233 (34% subtype B) and 1231 (16.5% subtype B) viral isolates. The levels of single base conservation were compared to a global consensus and calculated using Bioedit[13].

Identification of short continuous stretches of matching oligonucleotide sequences was carried out using the R statistical computing environment (version 2.8.1)[14] and tools from the seqinR package available at http://pbil.univ-lyon1.fr/software/SeqinR. Briefly, after removal of gaps, this implementation locates all possible substrings of a defined length within a reference sequence and calculates the corresponding frequencies of perfect matches to the alignment. Each amplicon was analyzed separately. The source code is available upon request.

Match frequencies (Table 1) of FH-probes were calculated based for either only B-subtypes or all HIV-1 groups (subtypes M, N, O, CPZ).

2.2. FH-probe design

All FH-probes contained the fluorophore FAM and the quencher TAMRA. Tms of the DNA FH-probes were calculated with an open source program (www.biophp.org) using the base-stacking algorithm, which was adjusted for the presence of 3mM $MgCl_2$. In all FH-probes a 5' terminal G was avoided[5,6,8]. The calculated Tms were aimed to exceed 55°C, except for mf341, which was designed solely based on phylogenetic considerations. DLNA FH-probes were designed with a length of 8-14 bases and with a content of LNA bases of 10-75%. Tm for DLNA FH-probes was calculated with an open source program different from that used for DNA FH-probes (http://lna-tm.com), considering differences in melting temperature between DNA and LNA bases.

2.3. Tm measurement

The Tm of the FH-probes was measured by examining the denaturation kinetics of the FH-probe (0.1uM) and its complementary strand (1uM) in a range from 30-90°C in 0.5°C intervals for 10 seconds using a real-time thermocycler (IQ5, Biorad, Basel, Switzerland). To attain conditions close to those of qPCR, the complementary strands were designed with overhangs of 3 bases avoiding dangling ends and the experiments were performed in 1 x Hotstartaq buffer (Qiagen) lacking enzyme with the inclusion of 3mM $MgCl_2$. The analysis of the melting curves was executed by subtracting the background fluorescence of the FH-probe without the complementary strand from the fluorescence signal in reactions containing the complementary target. The Tm was defined as the temperature at which 50% of subtracted fluorescence was reached.

2.4. PCR and signal to noise ratios

DNA qPCR was performed as described previously[15] using HotStarTaq Master Mix (Qiagen), 1uM of each primer and 0.1uM FH-probe. Experiments were done in duplicate using the real-time thermocycler IQ5 (BioRad, Basel) and as cycling profile: 95°C 15', 60x (95°C 5'', 55°C 5'', 60°C 40''). The following primers were used; cr1 (TCTCTGGCTAACTAGGGAACCCACTGCTT)[16] and cr2 (TGACTAAAAGGGTCTGAGGGATCTCTAGTTACCAG)[16] for the *early* region, ts5'gag (CAAGCAGCCATGCAAATGTTAAAAGA)[15] and skcc (TACTAGTAGTTCCTGCTATGTCACTTCC)[17] for the *gag* region, mf209 (AAAGCGTCTAGCCATGGCGTTAGTA) and mf302 (CAAATTTCTACTAATGCTTTTATTTTTTC) for the *pol* region, mf1 (CTTAGGCATCTCCTATGGCAGGAA)[18] and mf238 (GCTATTATTGCTGCTACTACTAATGCTACTA) for the *tat* region, mf222 (GGCAGGGATATTCACCATTATCGTTTCAGA) and mf83 (GGATCTGTCTCTGTCTCTCTCTCCACC)[18] for the *sa7* region, and mf345 (AATCAGGGAAGTAGCCTTGTGT) and mf346 (GAGGTGGGTTTTCCAGT) for the *nef* region . HXB2 was chosen as a standard target for all the experiments and pHXB2[19] was linearized by digestion with Xho1. Concentration of the plasmid was quantified by spectrophotometry. Based on the length of the linearized plasmid (13000bp), the molecular weight ($8.5x10^6$ g/mol) and copy number were calculated. A standard dilution series with 10fold dilution steps was prepared.

SNRs of individual FH-probes were determined by amplification of constant copy numbers ($3x10^6$ copies HXB2 DNA) dividing mean fluorescence of the last 5 cycles of the amplification when fluorescence has reached a plateau, by mean fluorescence of the initial 5 PCR cycles. Means of two independent experiments using duplicate measurements were calculated.

2.5. Calculations and Statistics

Statistical analyses were performed either using GraphPad Prism5.0 software (GraphPad Software, San Diego, CA) or Stata (Version 10.0; StataCorp).

Univariable and multivariable linear regression to determine predictors for SNRs of FH-probes were performed using Stata. $P<0.05$ was considered statistically significant. Factors with strong colinearity ($p<0.01$) were not included in the multivariable analysis.

The following parameters were tested in univariable analyses additionally to the ones shown in Table 2, but without revealing statistically significant association with SNRs: G/C-basepairs in 1st and 2nd position, purine in 1st and 3rd position, concentration of the FH-probe, Tm of the primer binding the opposite strand as the FH-probe and distance of FH-probe to the primer binding the same strand as the FH-probe.

3. Results

3.1. Phylogenetic complexity of conserved regions of the HIV-1 genome

Amplica in 6 regions commonly used for HIV-1 qPCR due to their conservation or usefulness for monitoring important splice variants were assessed in this study: An amplicon mapping the primary viral transcripts in the R/U5 region (*early*), the coding region for *p24-gag* (*gag*), the start of the reverse transcriptase gene (*pol*), the first coding region of the *tat*-gene (*tat*), *env-gp41* flanking the major splice acceptor 7 (*sa7*) and the poly-purine tract within the *nef* gene (*nef*) (Table 1, Fig. 1a).

Phylogenetic conservation within the six chosen amplica was analyzed with emphasis on match frequency, defined as 100% identity of FH-probe with its target, by employing a novel sequence scanning program. In this algorithm, a sequence defined as standard is screened in a sliding window approach for match frequencies of all possible FH-probes by comparison to a sequence database. In the present study, HXB2, the commonly used first molecular isolate of HIV-1[19], was used as the reference sequence and the database was a collection of 1242 near full-length HIV-1 genome sequences representing phylogenetically distinct viral isolates spanning the full scope of the HIV-1 pandemic, which was extracted from the Los Alamos HIV Sequence Database (hiv.lanl.gov). Figure 1b shows an analysis of 10-, 20- and 40-base oligomers plotted against average single base conservations. Theoretically, match frequencies (M) are proportional to the product of single base frequencies; this can be approximated by

equation 1) $M = (C_{GM})^{L}$,

which potentiates the geometric mean of single base conservation (C_{GM}) with oligonucleotide chain length (L).

A comparison of calculated versus observed match frequencies revealed close correlation (Pearson r=0.96) but significantly higher values for observed match frequencies (Wilcoxon signed rank test, $p<0.0001$). The most likely explanation for this discrepancy is that single base conservations were not independent from each other but genetically linked due to constraints of codon usage and RNA structure. Thus, to adjust the model for linkage, a correction factor (λ) was introduced:

Equation 2) $M = (C_{GM})^{L \times \lambda}$

Fitting the dataset to equation 2 resulted in λ=0.74 (95% confidence interval: 0.73-0.75). The model showed significant correlation (Pearson r=0.97) between observed

Table 1. Properties of FH-probes

ID[A]	FH-probe sequence[B]	amplica	position[C] 5'	3'	length	match (%)[D] all	nonB	B	SNR[E]	Tm[F] m	p
mf320	AAAGCTTGCCTTGAGTGCTTCA	early	530	551	22	36	29	52	4.4	64.5	67.7
mf321	TTCAAGTAGTGTGTGCCCGTCT	early	548	569	22	36	29	49	3.8	64.5	68.9
mf322	TGCCCGTCTGTTGTGTGACT	early	561	580	20	64	50	92	6.6	63.5	69.1
mf323	CCCGTCTGTTGTGTGACT	early	563	580	18	64	50	92	5.9	57.5	64.3
mf74	AGCACTCAAGGCAAGCTTTATTGAGGC [24]	early	548	522	27	87	82	97	3.9	64	72.1
ca20	TGtGtGCcCgT	early	557	567	11	83	77	96	12.9	59	68
ca21	TGtGtGccC	early	557	565	9	95	92	100	9.9	57	64
ca22	TGTGtGCCCGT	early	557	567	11	83	77	96	10.5	52	56
ca23	AGcTtGcCtTGaG	early	532	544	13	96	95	97	10.7	62	67
ca24	AGCttGCCttGaG	early	532	544	13	96	95	97	27.3	62	70
ca25	AGCtTGCCtTGAG	early	532	544	13	96	95	97	8.2	53	54
mf316	TAGAGTGCATCCAGTGCATGCA	gag	1428	1449	22	0.8	0	4.8	5.8	65	68.7
mf317	ATGCAGGGCCTATTGCACCA	gag	1445	1464	20	22	16	51	6.5	63	68.8
mf318	AGGCCAGATGAGAGAACCAAGG	gag	1464	1485	22	24	15	65	4	69	68.9
mf319	TGCAGCTTCCTCATTGATGGT	gag	1419	1399	21	43	36	78	4.8	63	66.8
Boe3	TCTATCCCATTCTGCAGCTTCCTCATT [25]	gag	1431	1405	27	15	3.3	70	1.9	70.5	70.6
Boe3.2	TCTATCCCATTCTGCAGCTTCCTCATTGATGG	gag	1431	1400	32	13	2.8	66	1.7	73	74.3
Boe3.3	TCCCATTCTGCAGCTTCCTCATTGATGG	gag	1427	1400	28	38	32	69	2	56.3	73.5
ri17	CCAtCaAtGaGGA	gag	1400	1412	13	68	64	88	5	56	59
ri18	CTGcAgAaTgGGA	gag	1415	1427	13	80	78	86	1.3	62	66
ri19	CATgCaGgGcCT	gag	1444	1455	12	59	57	70	4.2	63	72
ri20	AGGcCcTgc	gag	1455	1447	9	79	77	90	5.9	58.5	61
ri21	TCccattc	gag	1427	1420	8	87	87	89	1.2	nn	53
mf304	ATGGCCCAAAAGTTAAACAATGGCCA	pol	2599	2624	26	22	13	67	4.9	66	71
mf305	ATTCCTGGCTTTAATTTTACTGGTACAGT	pol	2596	2568	29	47	41	76	2.2	66.5	68.2
mf309	TTAAACAATGGCCATTGACAGAA	pol	2611	2633	23	63	60	75	2.3	57	64.2
mf348	AAGCCAGGAATGGATGGCC	pol	2586	2604	19	63	60	81	8.5	65.5	67
ri14	ACTgTaCcAgTA	pol	2568	2579	12	83	81	91	1.6	54	54
ri15	CAGgAaTgGaTGG	pol	2590	2602	13	87	86	95	9.5	59	62
ri16	CTGtCaAtGgCCA	pol	2631	2619	13	82	81	90	9.3	63.5	64
ri23	ATGgatgg	pol	2595	2602	8	97	96	98	1.5	53	54
mf82	TCTTCGTCGCTGTCTCCGCTTCTT [17]	tat	6001	5978	24	17	7.6	64	2.5	61.5	72.7
mf324	TCCCAACCCCGAGGGGA	sa7	8386	8402	17	2.7	0.8	12	2.8	64.5	69.8
mf325	CCCGAAGGAATAGAAGAAGA	sa7	8412	8431	20	5.4	5.9	2.9	3.9	58.5	61.2
mf326	CTCGGGGTTGGGAGGTG	sa7	8398	8382	17	0.6	0	3.4	2.9	63.5	66.6
mf327	CTATTCCTTCGGGCCTGTC	sa7	8424	8406	19	5.5	6	2.9	4.4	63.5	64.4
mf328	ACCCCGAGGGGACCC	sa7	8391	8405	15	2.7	1.1	11	4	62	67.2
mf2tq	TTCCTTCGGGCCTGTCGGGTCCC [17]	sa7	8421	8399	23	18	9.2	61	6.5	74.5	77.4
mf226	AGGGGACCCGACAGGCCC [15]	sa7	8397	8414	18	23	13	73	18.1	72	73.4
ri10	TGTcGgGtCcCCTC	sa7	8409	8396	14	32	24	74	3.8	73	75
ri11	CCTgTcgGGtCCC	sa7	8411	8399	13	38	29	79	3.2	59	73
ri12	CCCgACaGgCcCG	sa7	8403	8415	13	36	28	77	3.9	64	77
ri13	CGAcAgGcCCgAA	sa7	8405	8417	13	31	24	69	4.3	58	76
ri22	CGAcAggC	sa7	8405	8412	8	75	73	87	1.4	53	55
mf342	TTTTTAAAAGAAAAGGGGGGAC	nef	9064	9085	22	86	88	77	1.1	66.5	62.3
mf343	AAGAAAAGGGGGGACTGGA	nef	9071	9089	19	87	89	77	1.3	63	65.5
mf344	ACAGATCAAGGATATCTTGTCT	nef	9133	9112	22	8	2.4	36	1.7	54	61.4
mf341	CTTTTTAAAAGAAAAGGGG	nef	9063	9081	19	83	86	70	1	42	54.5
ca18	CAAGGCAGCTGTAGATCTTAGCCA	nef	9039	9062	24	3.4	0.5	18	3.8	63	68.8
ca19	AGGGCTAATTCACTCCCAACGA	nef	9090	9111	22	0.2	0.1	0.5	9.3	63	68.5
ca26	AGGGGgGaCtGgA	nef	9077	9089	13	95	95	95	4.2	58	74

[A] ID denominates identification of FH-probes according to internal nomenclature in the author's laboratory.
[B] Sequence of the FH-probe (5' to 3'): capital letters indicate DNA bases and lower case letters indicate LNA bases.
[C] Numbering based on the HXB2 genome according to the software provided by the Los Alamos HIV Sequence Database (SEQUENCE LOCATOR, www.hiv.lanl.gov).
[D] Conservation of all subtypes (all), nonB- (nonB) or B-subtypes (B) was calculated by comparison to sequences that were obtained from curated alignments in the Los Alamos HIV Sequence Database (download 2007, www.hiv.lanl.gov).
[E] Signal to noise ratio
[F] Measured (m) and predicted (p) melting temperature (Tm) (°C)

and predicted match frequency and no significant difference between them (Wilcoxon signed rank test, p=0.32). This analysis demonstrates that the probability that an oligonucleotide will find its match in a viral population decreases exponentially with FH-probe length (Fig. 1c). The influence of the oligonucleotide length may be relieved if genetic linkage results in coinheritance of clusters of bases, in which case, $\lambda<1$, oligonucleotide length is virtually shortened.

In contrast to these constraints related to assay sensitivity, specificity of FH-probe binding (S), defined as the probability that a certain sequence would not occur in a random amplicon, can be approximated by

equation 3) $S = (1-(1/4)^{L})^{(A-L+1)}$,

where L is the length of the FH-probe and A the length of a random, erroneously amplified sequence. This equation can be justified as follows: The probability, that the FH-probe matches a given stretch of length L (i.e. the length of the FH-probe) is $(1/4)^{L}$ (assuming equilibrium of all 4 bases). The entire amplicon can be viewed as consisting of A-L+1 stretches of length L. This approximation further assumes that the event of matching a given stretch is independent from matching the other stretches. Under this assumption the probability of obtaining no match at all is $(1-(1/4)^{L})^{(A-L+1)}$. Thus, it was estimated that the specificity of FH-probe binding exceeds 99% with a FH-probe length of $\geq$9bases and A ranging from 40-1000 bases (Fig. 1c).

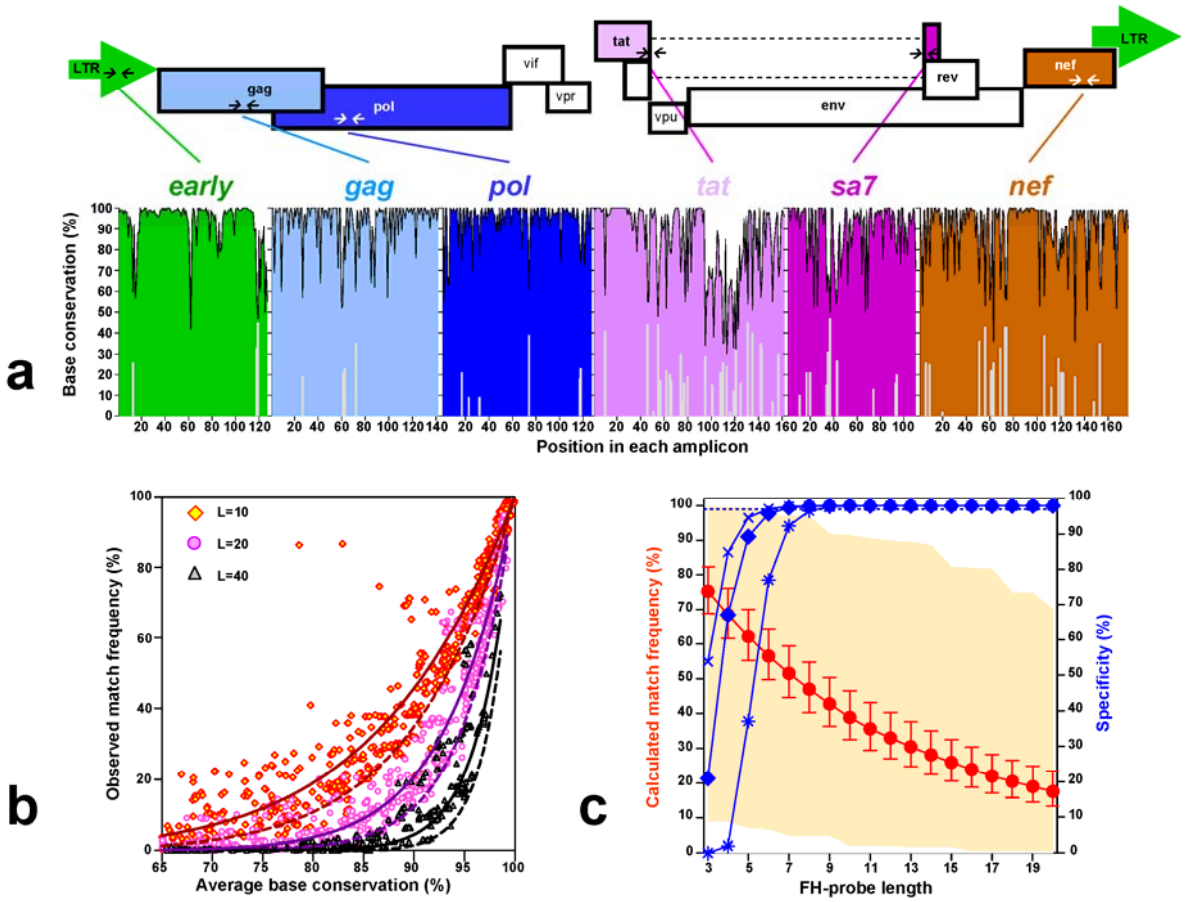

Figure 1. Phylogenetic complexity of conserved regions of the HIV-1 genome.(a) FH-probes mapping to 6 regions of the HIV-1 genome were designed utilizing the HXB2 sequence as a template. Upper panel: Small arrows mark the location of amplica in the viral genome. Lower panel: Conservation of the chosen amplica among all circulating HIV-1 clades (subtypes A-O, sequences obtained from curated alignments in the Los Alamos HIV Sequence Database, download 2007, www.hiv.lanl.gov). The colored area shows the level of single base conservation as compared to a global consensus sequence, and grey bars signify the conservation of the actual HXB2-sequence when this deviates from consensus. (b) Dependence of match frequencies of oligonucleotides on base conservation and oligonucleotide-length (L). HIV-1 sequence alignments were scanned for the presence of all possible 10- (red, yellow), 20- (magenta), and 40-mers (black) in the 6 chosen amplica. Match frequencies (M) were plotted against average (GM, geometric mean) of single base conservation (C). Broken lines show prediction of M by equation 1) $M = (C_{GM})^{L}$. Steady lines depict predictions by equation 2) $M = (C_{GM})^{Lx\,\lambda}$ including a correction factor (λ) to account for linkage. Brown lines show predictions for 10-mers, magenta lines for 20-mers and black lines for 40-mers. (c) Calculated match frequencies (M according to equation 2) (left y-axis; red circles, error bars: 95% confidence intervals, shaded area: range). Specificity (S, right y-axis) of FH-probe binding is modelled as a function of FH-probe length (L) and length of an erroneously produced amplicon (A): Equation 3) $S = (1-(1/4)^{L})^{(A-L+1)}$ (A=40 bases: blue crosses, A=100 bases: blue diamonds, A=1000 bases: blue asterisks)

3.2. Signal to noise ratios of FH-probes as indicators of DNA FH-probe performance

In previous studies signal to noise ratios (SNRs) of fluorescent probes were used to gauge efficacy of qPCR assays[20,21]. To confirm this concept, SNRs of individual FH-probes were calculated by dividing fluorescence at the end of the amplification by fluorescence during the initial cycles of amplification. To avoid interference of qPCR by possible primer dimer amplification and the resulting reduction of fluorescence in reactions with low copy numbers[22], high amounts of template ($3x10^6$ copies HXB2 DNA) were used. As shown in Figure 2a and Table 1, performance in a panel of 30 DNA-based FH-probes ranged from ineffective (mf342, SNR=1.1) to excellent (mf226, SNR=18.7). Cycle threshold (C_t) -values and their standard deviations, respectively, showed significant inverse correlation with SNRs (Pearson r=-0.52; p=0.004 and r=-0.59; p=0.0007 respectively), confirming association of SNRs with sensitivity and reproducibility of qPCR.

3.3. Predictors of SNRs of DNA FH-probes

To assess the factors influencing biochemical performance of the DNA FH-probes, SNRs and their different potential predictors were examined in univariable and multivariable linear regression models (Table 2). Melting temperature (Tm) showed weak but significant correlation to SNRs, either when measured or predicted by nearest neighbour analysis (p=0.033; measured and p=0.038; predicted). High G/C content of the DNA FH-probe was associated (p=0.002) with high SNRs. Furthermore, the presence of G/C rich secondary structures within the FH-probe was, in contradiction to current rules, a highly significant positive predictor of SNRs (p<0.001). FH-probes with intramolecular stem-loop structure containing one or more G/C base-pairs performed better than FH-probes with no or only a weak secondary structure (Fig. 2b, p=0.003). A trend for association of short FH-probe length (p=0.102) with high SNRs was observed, while Tm-difference of FH-probe and the primer binding the same strand (Delta-Tm) lacked correlation (p=0.742).

Upon encountering FH-probe bound to its template, taq polymerase can either remove the FH-probe by exonucleolytic cleavage or by strand displacement. We hypothesized that stability of the probe-template complex at its 5' end may favour exonucleolytic cleavage and generation of free, unquenched fluorophore. Thus,

thermal stability and base composition of the 5' ends of the FH-probes were analyzed in detail.

SNRs were positively associated with the predicted Tm of the ten 5' proximal bases (Pearson r=0.65, p=0.0001) (Fig. 2c). In agreement, FH-probes with a G or C at position 3 showed significantly elevated SNRs as compared to those with A or T (p=0.001, data not shown). At the second position, presence of G was favourable, followed by presence of A. Hence, SNRs of DNA FH-probes with a purine at position two were significantly higher than SNRs with a pyrimidine (Fig. 2d, p=0.01). No significant association of the bases at the first position with SNRs was observed.

Multivariable analysis, including statistically significant factors (p≤0.05 in univariable analysis), except for those with strong colinearity (p<0.01), resulted in a model in which 5'-proximal Tm (p=0.006), a purine at position 2 (p=0.035) and the numbers of intramolecular G/C-basepairs (p=0.004), explained 67% of the variability in SNRs (Table 2).

Thus, we suggest the following rules to design functional DNA FH-probes: 5'-proximal Tm ≥ 30°C, a purine at position 2 and ≥ 2 intramolecular G/C-basepairs.

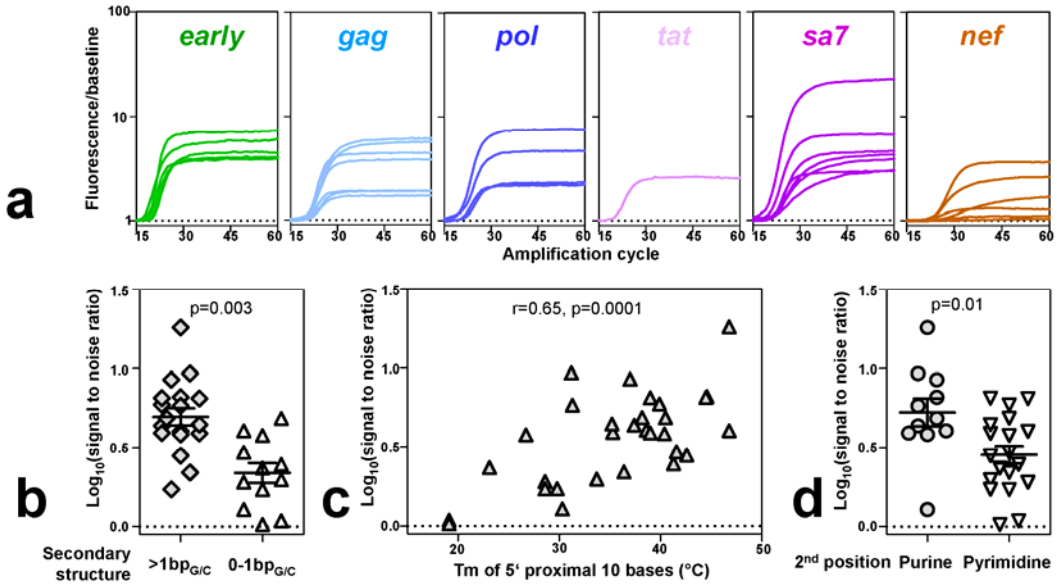

Figure 2. Performance of DNA FH-probes and analysis of predictors of SNRs. (a) Amplification curves of DNA FH-probes in 6 amplica (n=1-7, see Table 1). Fluorescence at each cycle was normalized to the baseline fluorescence of the first 5 cycles. Dotted lines indicate baseline fluorescence. (b) Comparison of SNRs of DNA FH-probes containing <1 (diamonds) or 0-1 G/C (triangles) base-pairs in a stem-loop. (c) Pearson correlation analysis of the Tm of the 5' proximal 10 bases of the DNA FH-probes with SNRs. (d) Comparison of DNA FH-probes containing a purine (circles) or a pyrimidine (triangles) at the 2nd position. P-values in (b) and (d) were calculated by Mann Whitney testing.

3.4. Performance of DNA FH-probes in qPCR

Four FH-probes with SNR>2 were chosen for further analysis according to their phylogenetic match frequencies: mf74 within *early* (87% match to viral isolates from all clades, 97.5% clade B), mf319 in *gag* (42% overall, 77.5% clade B), mf348 in *pol* (63% overall, 81% clade B), and mf226 in *sa7* (22.8% overall, 72.6% clade B). FH-probes within *nef* and *tat* were not included due to low match frequencies within the set of utilizable FH-probes (<20%). HIV-1 DNA copies were measured by qPCR in a range from 3 million to 0.3 copies per PCR and the 50% endpoint of PCR-positive dilutions was determined as an indicator of assay-sensitivity. All qPCRs using *gag*, *pol* and *sa7* FH-probes reached single copy sensitivity (Fig. 3b-d), while amplification of the *early* amplicon showed slightly reduced sensitivity presumably due to formation of primer dimers in later cycles (Fig. 3a), typically resulting in progressive diminution of signal amplitudes[15].

Thus, quantification of HIV-1 DNA was achieved over 6 orders of magnitude approaching single copy sensitivity for 3 FH-probes. Although they matched B-subtypes with reasonable frequency, match frequencies for nonB-subtypes were insufficient to cover the full scope of the global pandemic.

3.5. Improved short FH-probes using locked nucleic acid analogues

To design FH-probes with higher clade coverage, shorter chimeric DNA locked nucleic acid (DLNA) FH-probes were used. In locked nucleic acids base pairing is stabilized by a 2-O,4-C-methylene bridge on the ribose moiety[23]. As exemplified in Figure 4a, a 13-mer DLNA FH-probe showed higher thermal stability when bound to its complementary strand than an overlapping 23mer DNA FH-probe.

To identify candidate short FH-probes with high conservation, the sequence scanning algorithm was employed. An example is shown for the *pol* amplicon (Fig. 4b), which was scanned for all possible 8-, 10-, 12- and 14-mers. This procedure identified separate peaks of highly conserved sequence stretches from which FH-probes were chosen. Based on this analysis, 21 DLNA FH-probes (6 in *early*, 5 in *gag*, 4 in *pol*, 5 in *sa7*, 1 in *nef,* Table 1) ranging from 8-14 bases were synthesized.

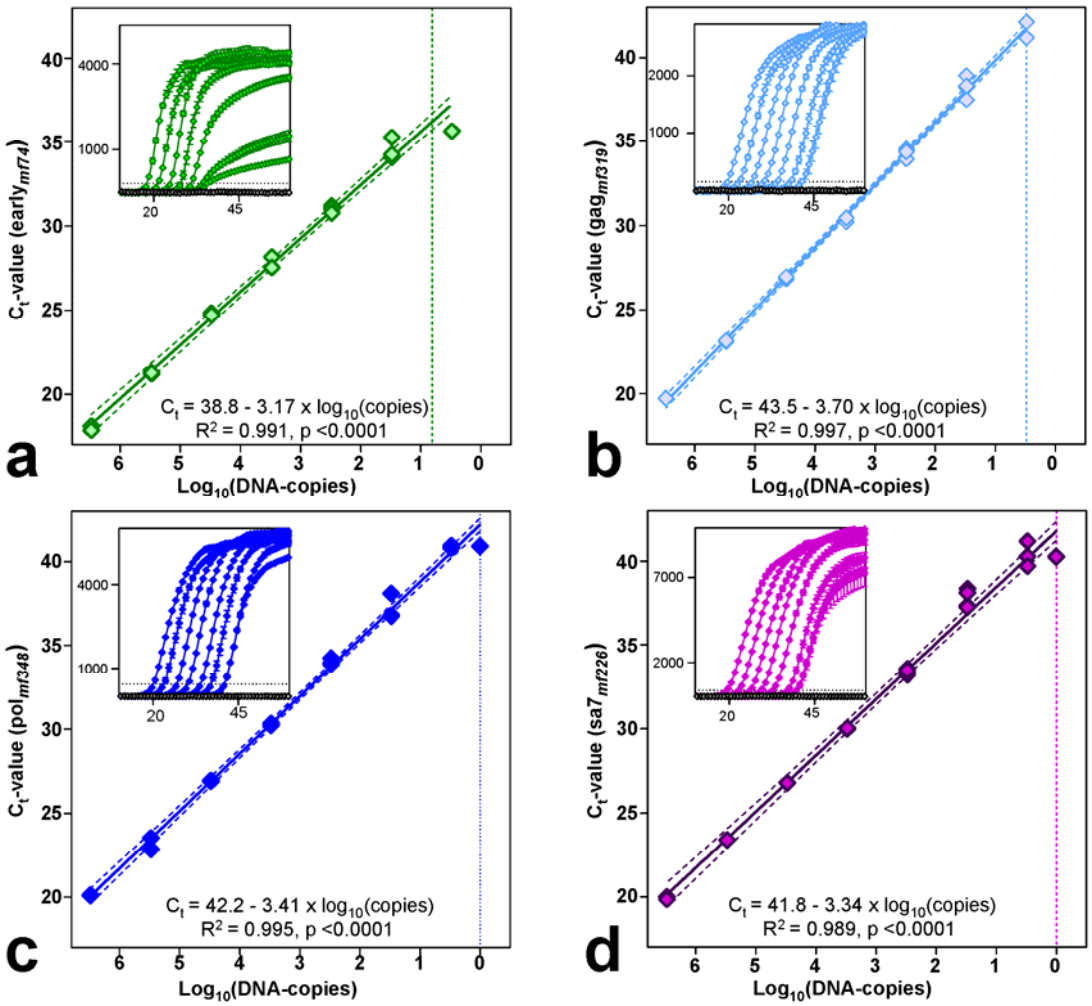

Figure 3. Sensitivity of qPCR using selected DNA FH-probes. DNA FH-probes with SNRs > 2 and the highest phylogenetic match frequencies of HIV-1 in each amplicon were used for amplification of HXB2 DNA in a range from 3 million to 0.3 copies per PCR. Broken vertical lines show the 50% endpoint of PCR-positive dilutions and thus signify sensitivity of the assays. qPCR for *gag* (b), *pol* (c) and *sa7* (d) reached single copy sensitivity, while amplification of the *early* (a) amplicon showed reduced sensitivity presumably due to primer dimer formation. Experiments were performed in duplicate ($3x10^6$-$3x10^3$ copies) or quadruplicate reactions (<$3x10^3$ copies). Results of linear regression of $\log_{10}$ of DNA copy numbers versus C_t-values are depicted. Inserts show amplification plots of PCR-cycle (x-axis) versus baseline subtracted fluorescence (y-axis), with dotted lines displaying the fluorescence threshold and grey symbols showing negative controls.

3.6. Predictors of SNRs of DLNA FH-probes

SNRs of DLNA FH-probes were measured as described for DNA FH-probes. C_t-values and their standard deviations, respectively, were inversely correlated with SNRs (Pearson r=-0.55; p=0.01 and Pearson r=-0.73; p=0.0003).

SNRs of DLNAs were tested for association with Tm, FH-probe length, G or C at position 3 and purine at position 2. Intramolecular G/C-basepairs were not assessed because DLNA FH-probes generally contained no secondary structures. Similarly, the Tm of the ten 5' proximal residues was not assessed as it was virtually identical to overall Tm. Finally, to account for the fact that not only G/C-residues promote strong basepairing but also LNA-residues, the presence of weakly pairing bases (plain A/T content, i.e. no G/C no LNA) was assessed rather than presence of strongly pairing bases (G/C, LNA).

The sole statistically significant predictor identified in univariable analysis was purine at position 2 (p=0.009). For identification of additional potential predictors, inclusion in the multivariable model was extended to parameters with a univariable p-value ≤0.1. This resulted in a model explaining 56% of the data variability (Table 2), in which purine at position 2 (p=0.007) and FH-probe length (p=0.023) positively predicted SNRs, while plain A/T-content showed a negative association (p=0.040).

Thus, we propose the following rules to design functional DLNA FH-probes: A purine at position 2, length ≥ 11 nucleotides and a plain A/T-content ≤ 35%.

Table 2. Predictors of SNRs of DNA and DLNA FH-probes

		Univariate analyses			Multivariate analyses			
		Coefficient (95% CI)		p	Coefficient (95% CI)		p	r^2
DNA (n=30)	Tm measured (°C)[A]	0.018	(0.001; 0.03)	0.038	0.005	(-0.01; 0.02)	0.356	0.68
	G/C content (%)[B]	0.013	(0.005; 0.02)	0.002				
	FH-probe length	-0.022	(-0.05; 0.005)	0.102				
	Predicted Tm 5' end (°C)[C]	0.025	(0.01; 0.04)	<0.0001	0.015	(0.005; 0.03)	0.006	
	G/C-basepair in 3rd position	0.331	(0.14; 0.52)	0.001				
	Purine in 2nd position	0.27	(0.07; 0.47)	0.01	0.155	(0.01; 0.30)	0.035	
	Delta-Tm (°C)[D]	0.003	(-0.01; 0.02)	0.742				
	Intramolecular G/C-basepairs	0.12	(0.06; 0.17)	<0.0001	0.074	(0.03; 0.12)	0.004	
DLNA (n=21)	Tm measured (°C)[A]	0.011	(-0.02; 0.05)	0.53				0.56
	Plain AT-content (%)[E]	-0.014	(-0.03; 0.002)	0.091	-0.014	(-0.03;-0.001)	0.04	
	FH-probe length	0.068	(-0.02; 0.15)	0.107	0.077	(0.01; 0.14)	0.023	
	G/C-basepair in 3rd position	0.044	(-0.31; 0.40)	0.797				
	Purine in 2nd position	0.423	(0.12; 0.73)	0.009	0.377	(0.12; 0.64)	0.007	
	Delta-Tm (°C) [D]	-0.009	(-0.03; 0.02)	0.453				

[A] Measured melting temperature.[B] Content (%) of G and C bases.[C] Predicted melting temperature of the 10 proximal bases at the 5' end. [D] Difference between the Tm of the FH-probe and the Tm of the primer binding the same strand as the FH-probe. [E] Content (%) of A and T bases, not including LNA bases

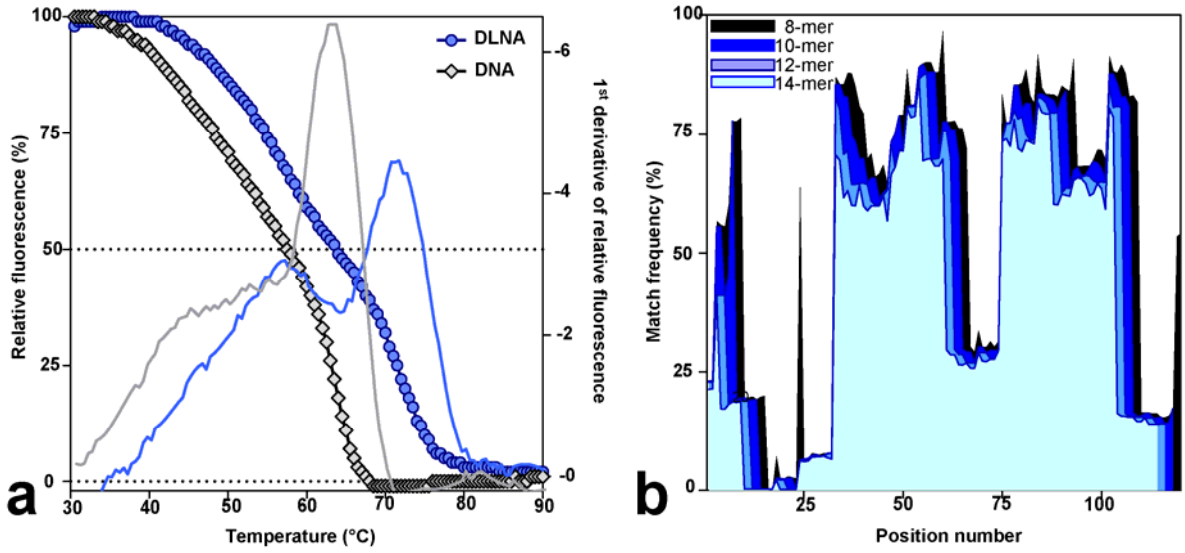

Figure 4. Melting profiles and match frequencies of DLNA FH-probes. (a) Melting profiles of two FH-probes within *pol*. The 13-mer DLNA FH-probe ri16 is depicted in blue and the overlapping 23-mer DNA FH-probe mf309 in grey. Symbols connected by lines show relative fluorescence and plain lines show their first derivatives. (b) Matching of HXB2 derived 8-, 10-, 12- and 14-mers to the aligned HIV-1 sequences of the Los Alamos HIV Sequence Database (download 2007, www.hiv.lanl.gov) starting at the indicated position within the *pol* amplicon (position 2500 to 2630, HIV-1 HXB2, GenBank accession number K03455).

3.7. Performance of DLNA FH-probes in qPCR

Based on their SNRs and phylogenetic conservation, DLNA FH-probes representing *gag (*ri20, 79% match to viral isolates from all clades, 90% in subtype B), *sa7* (ri12, overall 36%, subtype B 77%), *nef* (ca26, 95% overall and for clade B) and 2 FH-probes for *pol* (ri15, 87% overall, 94% clade B and ri16, 82% overall, 90% clade B) were further tested for their usefulness in qPCR (Fig. 5). FH-probes for the *early* amplicon were not included in this analysis because of the slightly reduced sensitivity of the qPCR, as shown above for DNA FH-probes, and because the dataset for this amplicon compassed a smaller number of viral variants than the other amplica (n=233 versus n=1242) with overrepresentation of subtype B (34% versus 17%). All PCR assays ranged over 6 orders of magnitude and reached single copy sensitivity.

In a previous study match frequencies of an RT-qPCR assay were boosted by adding two FH-probes into a single reaction[18]. The rationale was that fluorescence is expected to show up when at least one of two FH-probes A and B fits its target. This combined match frequency (M_{AvB}) can be estimated as

equation 4) $M_{AvB} = M_A + M_B - (M_A \times M_B)$.

This concept was employed with 2 FH-probes in *pol* (ri15 and ri16), which resulted in qPCR with single copy sensitivity (Fig. 5d). Assays performed with single FH-probes gave indistinguishable results (data not shown). Combinatorial analysis as described by equation 4, predicted that 97.8% (observed 96.7%) of viral isolates among all clades and 99.4% (observed 99.0%) within clade B, would match at least one of these two FH-probes. Thus, combination of two *pol* FH-probes resulted in near-universal clade coverage.

The *nef* DLNA FH-probe (ca26) showed the highest match frequency (95%) and good performance in DNA assays. However, when it was employed for RNA in one-step RT-qPCR under less stringent conditions (annealing at 50°C for 30 minutes), false positive reactions were observed at low frequencies presumably due to the stretch of 6 guanosines in a row. The other 4 DLNA FH-probes in this subset functioned reproducibly, with great specificity and without giving rise to false positive results using uninfected human PBMC DNA as a template, as well as in RT-qPCR (unpublished results).

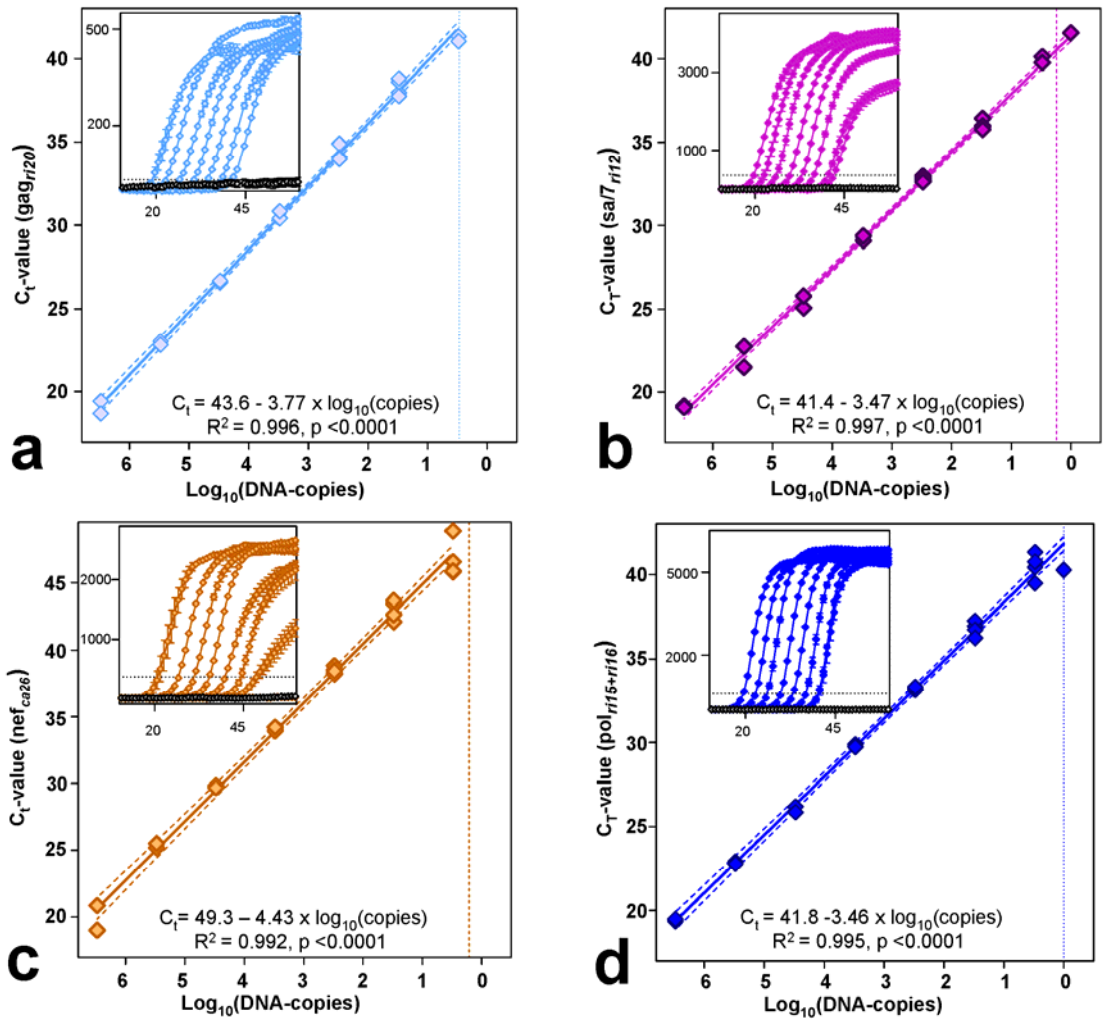

Figure 5. Sensitivity of qPCR using selected DLNA FH-probes. DLNA FH-probes with SNRs > 2 and the highest phylogenetic match frequencies of HIV-1 in each amplicon were used for amplification of HXB2 DNA in a range from 3 million to 0.3 copies per PCR. Broken vertical lines show 50% endpoints of PCR-positive dilutions and thus signify sensitivity of the assays. All qPCRs using *gag* (a), *sa7* (b), and *nef* (c) DLNA FH-probes reached single copy sensitivity. Two *pol* FH-probes that had both reached single copy sensitivity in qPCR on their own, were combined in one qPCR (d). This assay also reached single copy sensitivity. Experiments were performed in duplicate ($3x10^6$-$3x10^3$ copies) or quadruplicate reactions (<$3x10^3$ copies). Results of linear regression of $\log_{10}$ of DNA copy numbers versus C_t-values are depicted.
Inserts show amplification plots of PCR-cycle (x-axis) versus baseline subtracted fluorescence (y-axis), with dotted lines displaying the fluorescence threshold and grey symbols showing negative controls.

4. Discussion

In the present study, both the structural as well as the phylogenetic side of FH-probe design for HIV-1, a paradigm for a virus with high genetic diversity, were addressed. In consideration of the finding that match frequencies decrease exponentially with sequence length, usage of short optimally conserved FH-probes was aimed for. This has been facilitated by a novel computational tool to systematically identify stretches of highly conserved regions to which FH-probes may be targeted. This and the availability of locked nucleic acid analogues enabled us to design short FH-probes matching up to 95% of a representative collection HIV-1 isolates. Similarly, by combining two FH-probes within the same reaction, phylogenetic coverage could even be increased to 97%.

Regarding performance of FH-probes, analysis of the dataset in this study allowed to determine rules empirically for FH-probe design with the limitation that the collection of FH-probes was not strictly random and that certain rules were initially followed in design of FH-probes.

In agreement with current rules[6], a significant impact of stability of the probe-target complex both on DNA as well as on DLNA FH-probes was observed. However, more than the Tm, representing overall stability, the local stability near the site of nucleolytic cleavage of the FH-probe appeared to exert a marked influence on FH-probe performance. Hence, the Tm of the first 10 5'-proximal bases was a significant positive predictor of the SNRs. Moreover, stabilization of the probe-target duplex appeared to be mediated by forces beyond mere Watson Crick basepairing, as indicated by the observation that a purine at position 2 significantly predicted SNRs in the two independent datasets comprising DNA and DLNA FH-probes. This finding may reflect hydrophobic interaction of the FH-probe fluorophore with the extensive hydrophobic structure of the purine-rings.

Paradoxically, presence of intramolecular G/C base-pairs within DNA FH-probes was significantly associated with high SNRs despite their potential to compete with hybridization to the target strands. In analogy to molecular beacon probes which rely on stem-loops bringing fluorophore and quencher in juxtaposition[24], intramolecular G/C base-pairs in FH-probes may increase quenching in the absence of target. In agreement, significant correlation of intramolecular G/C base-pairs with baseline fluorescence (data not shown, Pearson r=-0.53, p=0.003) was observed.

Similarly to DNA FH-probes, the three identified predictors of DLNA FH-probe performance, purine 2, chain-length and low content of plain A/T base-pairs, all may influence duplex-stability. The fact that the Tm appeared not to predict FH-probe performance can be interpreted as a consequence of sampling, since DLNA FH-probes were chosen to exceed, whenever possible, a predicted Tm of 55°C in order to have a chance to perform well.

5. Conclusions

In summary, the following empirically tested positive predictors of biochemical FH-probe performance emerged from this study:

1. A purine at position 2 (DNA and DLNA).
2. An overall Tm ≥ 60°C and Tm of the first ten 5' proximal bases ≥ 30°C (DNA).
3. Presence of G/C rich (≥ 2 G/C-basepairs) secondary structures (DNA).
4. Sequence length (≥ 11) and low contents (≤ 35%) of plain A/T bases (DLNA).

Moreover, within the set of well performing FH-probes, application of the algorithm to scan sequence databases for FH-probes with optimal phylogenetic conservation, allowed to identify functional FH-probes in various regions of the HIV-1 genome approaching coverage of the global HIV-1 pandemic. It is conceivable that application of these phylogenetic and biochemical principles for FH-probe design may also be extended to other phylogenetically diverse biological systems.

Acknowledgements

This work was supported by a grant from the „Stiftung für Wissenschaftliche Forschung an der Universität Zürich" to MF and by the Swiss National Science Foundation (grant no. 3100A0-112670 to MF and grant no. 324730-116035 to HG).

We gratefully acknowledge Gertie Heimbeck (Biorad), for her skilful comments on real-time PCR and maintenance of real-time thermocyclers and Alexandra Trkola for her help in the initiation of the study and for helpful discussions.

References

1. Christopherson, C., Sninsky, J., and Kwok, S., The effects of internal primer-template mismatches on RT-PCR: HIV-1 model studies. Nucleic Acids Res 25 (3), 654 (1997).
2. Damond, F. et al., Human immunodeficiency virus type 1 (HIV-1) plasma load discrepancies between the Roche COBAS AMPLICOR HIV-1 MONITOR Version 1.5 and the Roche COBAS AmpliPrep/COBAS TaqMan HIV-1 assays. J Clin Microbiol 45 (10), 3436 (2007).
3. Ranade, K. et al., High-throughput genotyping with single nucleotide polymorphisms. Genome Res 11 (7), 1262 (2001).
4. Livak, K. J., Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genet Anal 14 (5-6), 143 (1999).
5. Gut, M., Leutenegger, C. M., Huder, J. B., Pedersen, N. C., and Lutz, H., One-tube fluorogenic reverse transcription-polymerase chain reaction for the quantitation of feline coronaviruses. J Virol Methods 77 (1), 37 (1999).
6. Mackay, I. M., Arden, K. E., and Nitsche, A., Real-time PCR in virology. Nucleic Acids Res 30 (6), 1292 (2002).
7. Bruijnesteijn Van Coppenraet, E. S. et al., Real-time PCR assay using fine-needle aspirates and tissue biopsy specimens for rapid diagnosis of mycobacterial lymphadenitis in children. J Clin Microbiol 42 (6), 2644 (2004).
8. Malnati, M. S. et al., A universal real-time PCR assay for the quantification of group-M HIV-1 proviral load. Nat Protoc 3 (7), 1240 (2008).
9. Livak, K. J., Flood, S. J., Marmaro, J., Giusti, W., and Deetz, K., Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. PCR Methods Appl 4 (6), 357 (1995).
10. Letertre, C., Perelle, S., Dilasser, F., Arar, K., and Fach, P., Evaluation of the performance of LNA and MGB probes in 5'-nuclease PCR assays. Mol Cell Probes 17 (6), 307 (2003).
11. Proudnikov, D. et al., Optimizing primer--probe design for fluorescent PCR. J Neurosci Methods 123 (1), 31 (2003).
12. Lunge, V. R., Miller, B. J., Livak, K. J., and Batt, C. A., Factors affecting the performance of 5' nuclease PCR assays for Listeria monocytogenes detection. J Microbiol Methods 51 (3), 361 (2002).
13. Hall, T. A., Bioedit: A user-friendly biological sequence alignment editor and analysis program for Windows 95=98 NT. Nucleic Acids Symp Ser 41, 95 (1999).
14. Ihaka, R. and Gentleman, R., R: a language for data analysis and graphics. Journal of computational and graphical statistics, 299 (1996).
15. Kaiser, P. et al., Productive Human Immunodeficiency Virus 1 Infection in peripheral Blood Predominantely Takes Place in CD4/CD8 double negative T Lymphocytes. J Virol 81, 9693 (2007).
16. Lewin, S. R. et al., Use of Real-Time PCR and molecular beacons to detect virus replication in human immunodeficiency virus type 1-infected individuals on prolonged effective antiretroviral therapy. J.Virol. 73, 6099 (1999).
17. Christopherson, C. et al., PCR-Based assay to quantify human immunodeficiency virus type 1 DNA in peripheral blood mononuclear cells. J.Clin.Microbiol. 38 (2), 630 (2000).

18. Fischer, M. et al., Cellular viral rebound after cessation of potent antiretroviral therapy predicted by levels of multiply spliced HIV-1 RNA encoding nef. J Infect Dis 190 (11), 1979 (2004).
19. Ratner, L. et al., Complete Nucleotide-Sequences of Functional Clones of the Aids Virus. Aids Research and Human Retroviruses 3 (1), 57 (1987).
20. Yao, Y., Nellaker, C., and Karlsson, H., Evaluation of minor groove binding probe and Taqman probe PCR assays: Influence of mismatches and template complexity on quantification. Mol Cell Probes 20 (5), 311 (2006).
21. Kutyavin, I. V. et al., 3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. Nucleic Acids Res 28 (2), 655 (2000).
22. Kaiser, P. et al., Equal amounts of intracellular and virion-enclosed hepatitis C virus RNA are associated with peripheral-blood mononuclear cells in vivo. J Infect Dis 194 (12), 1713 (2006).
23. Jensen, G. A., Singh, S. K., Kumar, R., Wengel, J., and Jacobsen, J. P., A comparison of the solution structures of an LNA : DNA duplex and the unmodified DNA : DNA duplex. Journal of the Chemical Society-Perkin Transactions 2 (7), 1224 (2001).
24. Tyagi, S. and Kramer, F. R., Molecular beacons: probes that fluoresce upon hybridization. Nat Biotechnol 14 (3), 303 (1996).

2. Novel method to select and clone HIV-1 derived small noncoding RNA of low abundance

Manuscript submitted as Althaus *et al*.

Own contributions

My major contributions to this study were as follows: I performed experiments to establish and apply the hybridization capture method for the enrichment of sncRNAs. I established HIV-1 sncRNA specific PCRs to investigate the presence of HIV-1 derived sncRNAs in infection experiments with the primary virus isolates, which I used to infect PBMC from different HIV-1 negative blood donors. Furthermore, I sequenced some exemplary PCR products to confirm the presence of HIV-1 sncRNAs in various virus isolates. I processed all raw sequence data of all libraries and analyzed the resulting small noncoding RNAs (sncRNAs). In particular, this included the performance of alignments to determine sequences with >90% homology to HIV-1, and the determination of unique HIV-1 sncRNAs which are derived from a different library, and/or are differing in length and/or position. I also carried out the analysis of the non HIV-1 clones to group them to different classes of cellular small noncoding RNAs. Additionally, I characterized the single unique HIV-1 sncRNAs according to their length and secondary structure. I performed all statistical analyses and wrote the manuscript.

Novel targeted enrichment strategy enables detection of low abundant small noncoding RNAs in HIV-1 infection

Claudia F. Althaus[1,3], Valentina Vongrad[1,3], Barbara Niederöst[1], Beda Joos[1], Francesca Di Giallonardo[1], Philip Rieder[1], Jovan Pavlovic[2], Alexandra Trkola[2], Huldrych F. Günthard[1,4], Karin J. Metzner[1,4] & Marek Fischer[1,4]

[1]Division of Infectious Diseases, University Hospital Zurich, University of Zurich, Switzerland; [2]Institute of Medical Virology, University of Zurich, Switzerland; [3]These authors contributed equally to this work; [4]These authors also contributed equally to this work.

This article is dedicated to the memory of Marek Fischer, who died in December 2010.

The various classes of small noncoding RNAs (sncRNAs) are important regulators of gene expression. While a rapidly increasing number of sncRNAs has been identified over recent years, isolation of sncRNAs of low abundance remains challenging. Virally encoded sncRNAs can be expressed at very low levels. This is best illustrated by HIV-1 where virus encoded sncRNAs represent at the outmost 0.1-0.5% of all sncRNAs. Here we report on a novel, sequence targeted enrichment strategy which allows a greater than 100-fold enrichment of low abundant sncRNAs. Using HIV-1 as model system, we generated sncRNA libraries from HIV-1 infected cells, which led to 892 individual HIV-1 sncRNA clones representing more than 70% of all sncRNA clones identified. This enormous capacity to enrich low abundance sncRNAs in a sequence specific manner highly recommends our selection strategy for any type of investigation where origin or target sequence of the sought-after sncRNAs are known.

Introduction

One major posttranscriptional regulatory pathway, RNA interference (RNAi), is mediated by small noncoding RNAs (sncRNAs)[1]. Over recent years, the importance of the diverse classes of sncRNAs has been widely recognized and their impact on various biological processes demonstrated across a broad variety of organisms[2]. The most intensively studied class of sncRNAs are the 20-25 nucleotides long microRNAs (miRNAs) which play a crucial role in posttranscriptional regulation of gene expression[3].

Despite technological advances sncRNAs of low abundance have remained difficult to identify. Until to date, the most frequently employed method to derive sncRNAs is the generation of cDNA libraries encoding sncRNAs by, rather rate limiting, cloning and sequencing procedures[4]. More recently, high-throughput sequencing techniques have been applied[5,6]. While these techniques allow the identification of sncRNAs of medium to high frequency with notable success, they remain less effective in defining low abundant sncRNAs. Alternate approaches have employed microarray- and PCR-based technologies to detect and quantify sncRNAs[4,7]. However, due to the short length of oligonucleotides used in microarrays and the target specificity of PCR, these procedures only lend themselves towards analyses where already known or predicted sncRNAs need to be detected.

Discovery and screening for viral sncRNAs in infected cells faces two challenges: Firstly, sequence and length of these viral sncRNAs are yet unknown excluding approaches which depend on target specific amplification. Secondly, depending on the virus studied, virus-encoded sncRNAs may be of extremely low abundance. The first discovery of viral miRNAs was made in Epstein-Barr virus (EBV)-infected human cell lines[8] where 4.15% sncRNAs of EBV origin were identified. The specificity could be enhanced by employing subtractive hybridization which yielded libraries consisting of ~40% EBV derived sncRNAs[9]. A similar high abundance of viral sncRNAs was also observed in cells infected with other DNA viruses[10]. However, sncRNAs from RNA viruses have thus far proven less frequent, accounting commonly for <1% of all sncRNAs in infected cells[11].

HIV-1 generates sncRNAs in very low abundance[10,12], if at all[13]. So far, only four sncRNAs with miRNA-like functionality have been identified in HIV-1 infected cells and mapped to domains in TAR[14,15], env[16], nef[17] and U3[18]. The first published report on screening for sncRNAs in HIV-1 infected cells detected only two viral sncRNAs in 1,540 clones from HIV-1 infected HeLa $T4^+$ cells (0.13%). No functional property could be assigned to these HIV-1 sncRNAs and they were accordingly classified as degradation products by the authors[10]. Another study screened 600 sncRNA clones derived from HIV-1 infected cells for HIV-1 sncRNAs but found none which contained a viral sequence[13]. More recently, two independent surveys performed high-throughput sequencing of HIV-1 infected cell libraries and detected 0.26%[12] and 0.50% (Schopman, N., Bradley, T., Willemse, M., van Kampen, A., Baas, F., Berkhout, B., Haasnoot, J., Keystone Symposia, Whistler, Canada, 20-25 March 2011, abstract #418) HIV-1 sncRNAs in approximately 48,000 and 5.2 million screened sncRNAs, respectively.

As these studies highlight, identification of low abundant sncRNAs, such as HIV-1 encoded sncRNAs, requires either screening of a large number of sequences or an optimized selection protocol. Here we report on a novel selection and enrichment strategy for low abundant sncRNAs. Key to this approach is a highly effective enrichment by hybridization capture where hybridization probes covering the entire genome of the organism of interest - in our case HIV-1 - are included. This approach is highly successful in detecting low abundant HIV-1 sncRNAs in cDNA libraries obtained from HIV-1 infected primary human cells. The yield of HIV-1 sncRNAs increased from previously reported 0.1-0.5% to an average of 78.3% (±7.6% (SD)) of total sncRNAs in several independent libraries allowing for the first time a non-biased evaluation of the regulatory properties of HIV-1 sncRNAs during the course of the infection.

Our novel approach to enrich for low abundant sncRNAs is highly effective and can easily be adapted to other model organisms or genomic regions of interests where an analysis of the complete spectrum of low and high abundant sncRNAs is warranted for.

Results

Enrichment and selection of low abundant HIV-1 sncRNAs by hybridization capture

The overall aim of our study was to derive an effective selection strategy for low abundant sncRNAs which would allow 1) to determine the presence or absence of sncRNAs in a given setting and 2) allow the characterization of the full spectrum of sncRNAs generated by a specific organism. Our model organism was HIV-1 where conflicting reports have been published which suggested that either no or only extremely low numbers of HIV-1 sncRNAs are evolved in infected cells. Our intent was foremost to derive a protocol that allows sufficient enrichment of low abundant sncRNAs. As outlined in the following, we achieved this by introducing a specific selection step which enriched for HIV-1 derived sequences. Figure 1 illustrates the various steps involved in our sncRNA selection procedure. To generate sncRNA libraries we first infected human primary cells (either $CD4^+$ T-lymphocytes or macrophages) with the HIV-1 isolate JR-FL (Fig. 1, Step 1). Seven to 14 days post infection, when a sufficient degree of infection of the cultures was reached as judged by p24 antigen production, small RNA species (<200 nt) were extracted from the infected cultures using a commercially available extraction procedure (Fig. 1, Step 2). Derived small RNAs where then subjected to C-tailing of the 3'-end and adaptor ligation at the 5'-end (Fig. 1, Step 3). To generate a cDNA library, RNA was reverse transcribed and cDNA amplified by PCR (Fig.1, Step 4). The next step was key for the success of our procedure as we here enriched for HIV-1 encoded sncRNAs by specifically selecting HIV-1 sncRNAs which bound to single-stranded HIV-1 DNA in a hybridization step (Fig.1, Step 5). The HIV-1 ssDNA hybridization probes used for this purpose were generated from proviral DNA of $HIV\text{-}1_{JR\text{-}FL}$ by PCR. In total, five probes covering the entire HIV-1 genome were generated (Fig. 1, Box 1). The primers used to amplify those hybridization probes were biotinylated which allowed to couple the derived probes to streptavidin beads. Adaptor-ligated cDNA derived in Step 4 was then hybridized to the HIV-1 ssDNA hybridization probes, followed by a magnetic bead purification step to eliminate nonhybridized cDNA species (Fig. 1, Step 5). The five HIV-1 ssDNA hybridization probes were either used together (as shown in Fig.1, Step 5) or in separate reactions. Both approaches proved equally

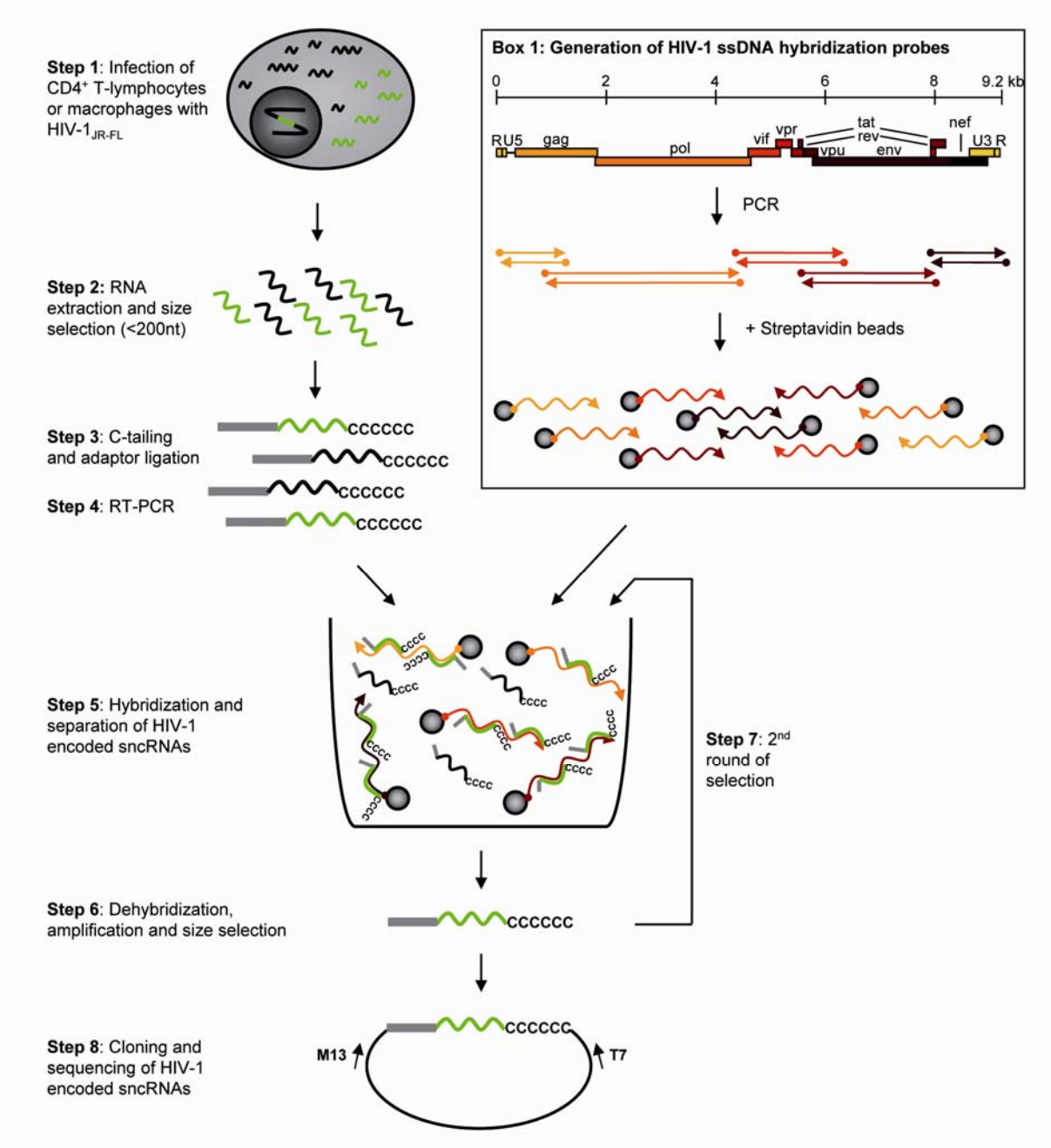

Figure 1. Scheme of cDNA library generation with hybridization capture for HIV-1 encoded small noncoding RNAs (sncRNAs). Primary human cells (macrophages or $CD4^+$ T-Lymphocytes) were infected with HIV-1_{JR-FL} (Step 1). Cellular (black) and HIV-1 encoded (bright green) sncRNAs (<200 nt) were extracted from primary HIV-1 infected cells (Step 2). RNA was C-tailed at the 3'-end, adaptor-ligated at the 5'-end (Step 3) and RT-PCR was performed (Step 4). For the preparation of the HIV-1 ssDNA hybridization probes, PCR was performed with biotinylated primers for 5 overlapping regions of the genome using HIV-1_{JR-FL} plasmid as template. Biotinylated amplicons were attached to streptavidin beads (Box 1). Sequences homologous to HIV-1 were enriched by incubation of cDNA derived from adaptor-ligated sncRNAs with a mixture of the 5 different HIV-1 ssDNA hybridization probes (Step 5); alternatively each HIV-1 ssDNA hybridization probe was used separately. After hybridization capture, bound amplicons were eluted, amplified and size selected on a gel (Step 6). Either the hybridization and size selection steps were repeated (Step 7) or the amplicons were directly cloned and sequenced after one round of selection (Step 8).

effective. Bead enriched cDNA was then dehybridized, amplified by PCR and additionally size selected (Fig. 1, Step 6) before undergoing either a further round of hybridization selection (Fig. 1, Step 7) or cloning into vector DNA and sequence analysis (Fig. 1, Step 8).

We successfully employed this procedure, performing one round of selection, for two independent cDNA libraries which yielded 4.8% and 12.9% clones with sequence homology to HIV-1 (Fig. 2a, Supplementary Table 1), respectively. While the achieved enrichment for HIV-1 sncRNAs was already more than an order of magnitude higher than frequencies reported in the previously published studies, we aimed to further enrich HIV-1 sncRNAs by performing a second round of hybridization capture. We generated in total seven sncRNA libraries that underwent two consecutive hybridization selections and were all highly enriched for HIV-1 sncRNAs yielding on average 78.3% (±7.6% (SD)) HIV-1 encoded clones (Fig. 2a, Supplementary Table 1). These results highlight that our approach has a striking capacity to enhance retrieval of low abundant sncRNAs. In our model system we achieved a greater than 100-fold increase in the selection of HIV-1 encoded sncRNA species over levels reported in the literature.

For four libraries from two independent experiments (libraries F-J, Fig. 2a) we separated the dehybridized cDNA into two fractions of 50-80 and 80-110 base pairs length, which, after subtracting the lengths of adaptors and the C-tail, leads to lengths of HIV-1 sncRNAs of ≤25 and 25-55 bp, respectively, before subjecting the cDNA to a second round of hybridization enrichment. With this approach we wanted to explore if the target molecule length has an influence on hybridization efficacy. The latter was a reasonable concern as it was previously suggested that short molecules are difficult to select by hybridization capture[19]. However, we could not confirm this in our setup. While as expected the separate size selection resulted in a significant difference of the median size of sncRNAs (25 nt [interquartile range, IQR: 22-33] and 44 nt [IQR: 33-51] for the 50-80 and 80-110 bp fraction, respectively, $p<0.0001$, Wilcoxon rank sum test) (Fig. 2b), the specificity of the hybridization capture for the smaller size sncRNA fraction was only slightly lower than for the larger size fraction (69.5% vs. 81.3% HIV-1 sncRNAs). 146 of 364 (40.1%) sncRNA clones showed a length of 20-25 nucleotides in the smaller size fraction as compared to 41 of 386

(11.1%) in the larger size fraction (p<0.0001, Chi square test). We can safely conclude that also sncRNA clones of smaller size can be efficiently derived using our hybridization capture.

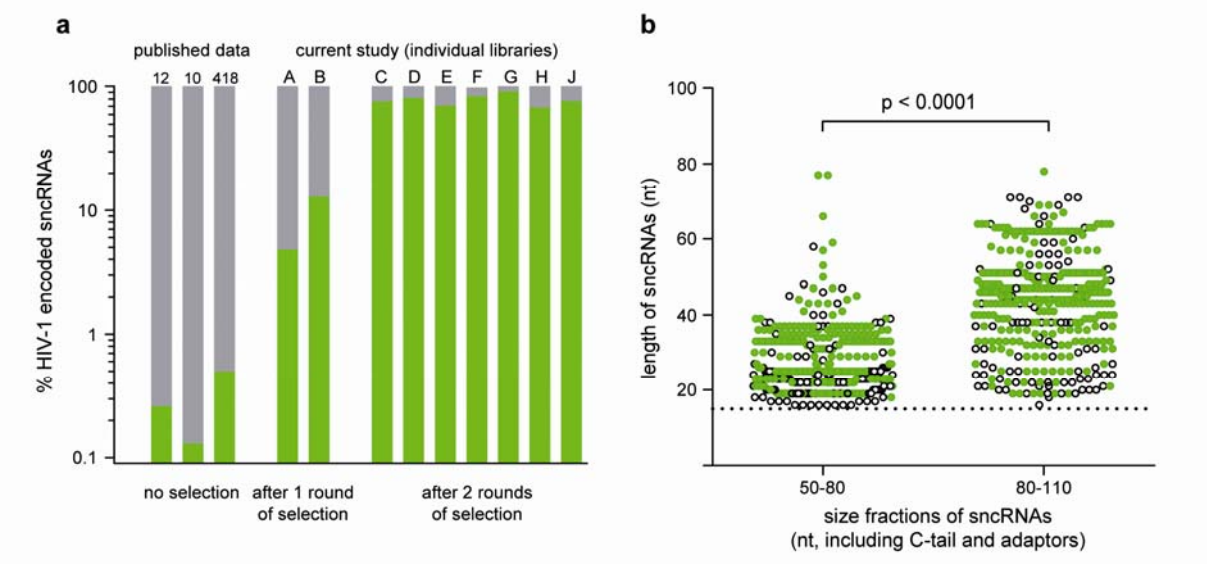

Figure 2. Efficiency of hybridization capture to enrich HIV-1 encoded sncRNAs. (a) Comparison of published data using the currents standard protocols (left) with our novel selection strategy (right). Standard protocols with no selection led to a yield of 0.1-0.5% HIV-1 sncRNAs[10,12] (and Schopman, N., Bradley, T., Willemse, M., van Kampen, A., Baas, F., Berkhout, B., Haasnoot, J., Keystone Symposia, Whistler, Canada, 20-25 March 2011, abstract #418). Numbers above bars indicate the respective literature references. Using our novel method, application of one round of hybridization capture yielded 8.9±5.7% (mean±SD) HIV-1 sncRNAs (libraries A and B). Performing two consecutive rounds of selection (libraries C-J) optimized the yield to 78.3±7.6% (mean±SD) HIV-1 sncRNAs. (b) Probing the influence of target molecule length on hybridization efficacy. Libraries F, G, H and J underwent a second size separation step before undergoing a second round of hybridization enrichment. Dehybridized cDNA was separated into two fractions of 80-110 bp or 50-80 bp length and both probed separately for hybridization efficacy. Green full circles denote HIV-1 derived sncRNAs, black open circles denote non HIV-1 sncRNAs. Both fractions successfully retrieved sncRNA in the second round hybridization. 20-25 nucleotide long sncRNAs were retrieved from both fractions and comprised 40.1% of all sncRNAs in the small size fraction and 11.1% in the large size fraction (p<0.0001, Chi square test).

To verify that the individual HIV-1 ssDNA hybridization probes used select specifically HIV-1 sncRNAs of the respective region, we generated two libraries (H and J) where HIV-1 ssDNA hybridization probes were utilized in separate reactions in the two rounds of selection. We found that 92.8±7.9% (mean±SD) of the thereby recovered HIV-1 sncRNAs were specifically enriched (Supplementary Table 2). Interestingly, HIV-1 sncRNAs within the RU5 region (contig 2, Supplementary Table

3) could be selected with each HIV-1 ssDNA hybridization probe, regardless of the specific sequence. The latter possibly indicates a higher abundance within the HIV-1 sncRNA species, or a sequence homology-independent binding modus to the HIV-1 RNA genome.

Characterization of HIV-1 small noncoding RNAs

In total, we derived 1,335 clones from nine individual sncRNA libraries generated from HIV-1 infected primary cells after one or two rounds of hybridization capture (A-J, Supplementary Table 1). Clones were defined as valid sncRNA candidates when they 1) contained the C-tail and the 3' and 5' adaptor sequences and 2) were in the expected size range (>15 and <100 nt). 892 of these clones had a greater than 90% homology to the strain HIV-1_{JR-FL} used for infection. Of these, 216 clones were distinguishable as unique clones by various measures (e.g., derived from different libraries or differed in length and/or position; Supplementary Table 3). It can be reasoned that identical clones within one library may indicate sncRNA species which occur at higher abundance. However, deriving quantitative conclusions from our type of analysis is difficult as it cannot be ruled out that preferential amplification of certain clones occurred during PCR.

We aligned the 216 unique HIV-1 sncRNAs to the reference strain HIV-1_{HXB2} (Fig. 3a, Supplementary Table 3). They had a length of 43±14 nucleotides (mean±SD, range: 16-89 nt) (Fig. 3b). Based on this alignment we found that the derived HIV-1 sncRNAs grouped within 67 different contigs, i.e., single or clusters of overlapping HIV-1 sncRNAs. 45 contigs (67.2%) contained 2 to 17 unique sncRNAs; 37 contigs harbored sncRNAs identified in at least two different libraries highlighting that these sncRNAs were not formed randomly. The contigs were spread throughout the HIV-1 genome and the majority of them consisted entirely of sense sncRNAs (56 contigs, 84%). 21 antisense sncRNAs were detected in either antisense only contigs (6 contigs, 9%) or in mixed sense and antisense contigs (5 contigs, 7%). Of note, sncRNAs with differential polarity in these mixed contigs have the potential to form double-stranded sncRNAs.

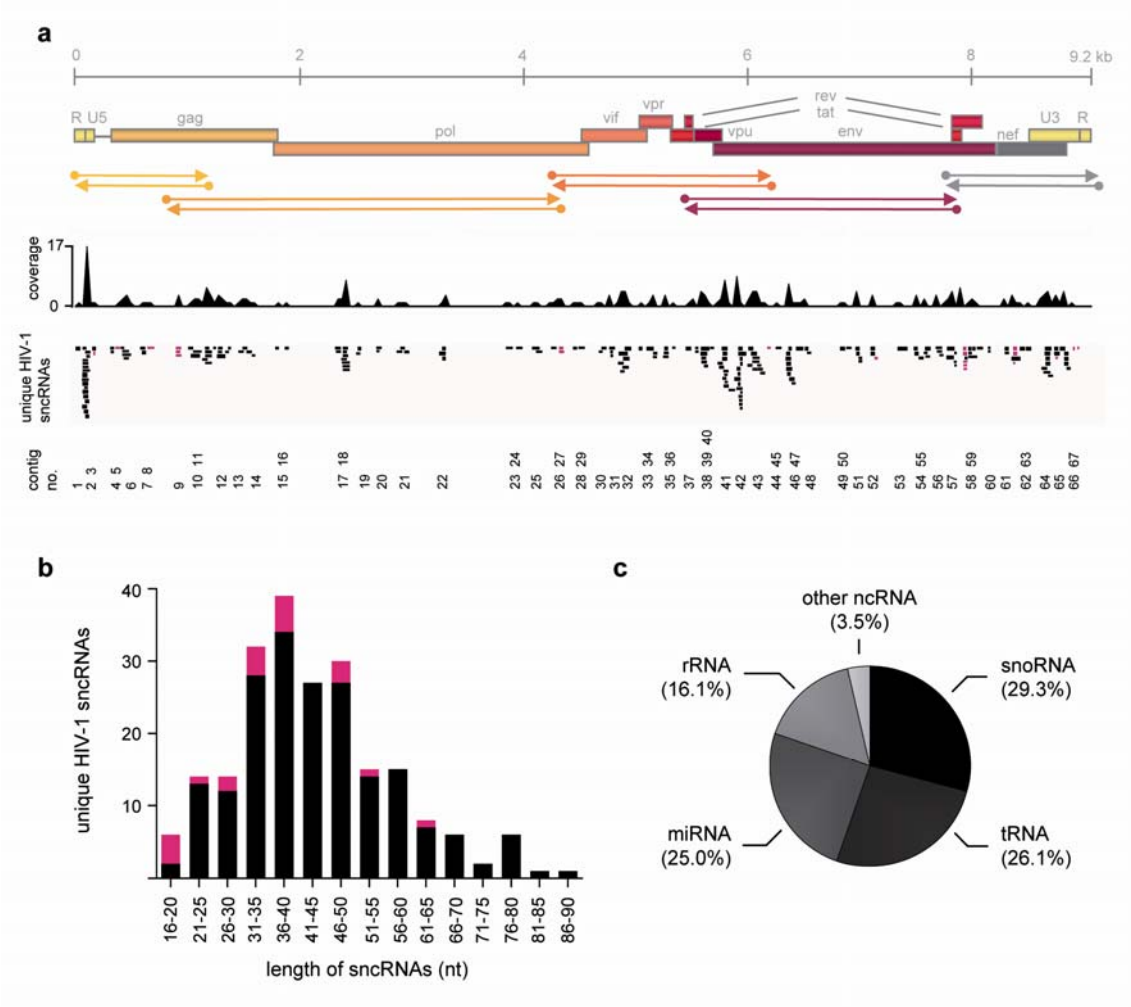

Figure 3. Alignment and characterization of identified sncRNAs. (a) 216 unique sncRNAs from nine libraries (Supplementary Table 3) were aligned to the reference strain HIV-1_{HXB2} and cluster in 67 contigs distributed throughout the whole genome of HIV-1. The coverage of sncRNA per nucleotide, the single unique clones (black: sense HIV-1 sncRNAs, pink: antisense HIV-1 sncRNAs), and the contig numbers are shown. (b) The length (nucleotides, nt) distribution of all unique HIV-1 sncRNAs is depicted. Sense sncRNAs are shown in black, antisense sncRNAs are shown in pink. (c) Pie chart depicting the distribution of different human cellular sncRNAs in all libraries.

Our selection procedure was highly successful in both, selecting a high number of HIV-1 sncRNAs but also in defining new HIV-1 sncRNA species. Of the identified 216 unique HIV-1 sncRNAs, eight correspond to previously described HIV-1 miRNAs: Six sncRNAs correspond to hiv1-miR-N367[17] within nef, one to hiv1-miR-TAR-3p[15] and one to hiv1-miR-H1[18]. Of particular note, while not identical in sequence and length, approximately 70% of our HIV-1 sncRNAs overlap with the 125 HIV-1 sncRNAs detected by pyrosequencing[12].

Although our selection strategy for HIV-1 encoded sncRNAs is highly effective, we still retrieved approximately 30% of sncRNAs which were not of HIV-1, but most of them of human origin. Other clones contained plasmid-derived, bacterial or unknown sequences, i.e., sequences without a match in the GenBank database. The majority of those human sequences (86%) could be assigned to various classes of human cellular sncRNAs, i.e., miRNAs, small nucleolar RNAs and transfer RNAs (Fig. 3c). As expected, $tRNA_{Lys}$ was frequently found since this tRNA functions as primer for the initiation of reverse transcription of the HIV-1 RNA.

HIV-1 sncRNA contigs identify regions for sncRNA generation across different exemplary HIV-1 primary virus isolates

While we were successful in demonstrating that sncRNAs are generated in HIV-1 infected cells, our analysis was based on the virus isolate JR-FL. We thus sought to explore whether the identified sncRNAs are specific for this particular virus or are ubiquitously generated in HIV-1 infection. As proof-of-principle, we investigated the presence of three sncRNA contigs (contig 2, located in the conserved LTR region, and contigs 43 and 58 both located in env, Supplementary Table 3) in PBMC from HIV-1 uninfected donors infected with five unrelated patient-derived primary virus isolates, which were obtained during acute HIV-1 infection[20]. Isolates were chosen based on the patient-specific sequences of the env gene (Philip Rieder, Beda Joos, unpublished data) to assure annealing of the specific sncRNA primers. By specifically tailored RT-PCR all three HIV-1 sncRNA contigs were detected in most of the infected cultures (Fig. 4a), indicating that these sncRNAs are not specific for an individual virus strain, nor produced randomly as they emerge upon infection with genetically divergent HIV-1 strains.

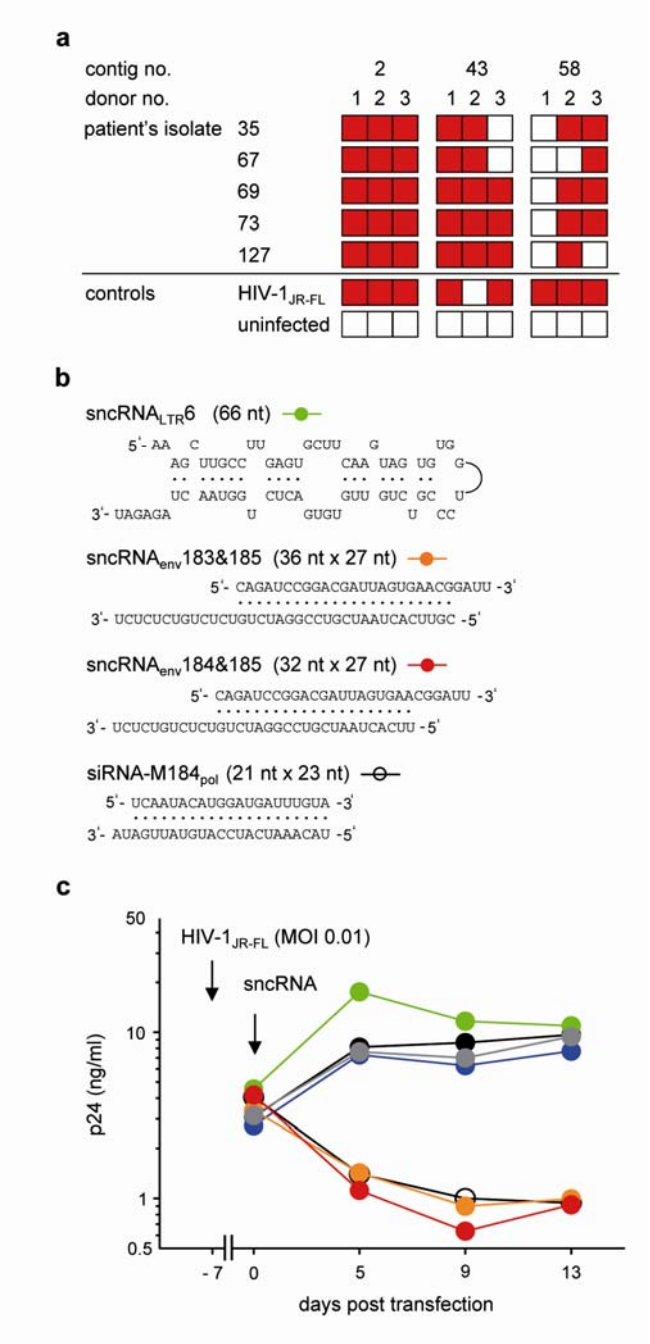

Figure 4. Functional analysis of HIV-1 sncRNAs. (a) $CD4^+$ T-lymphocytes of three healthy donors were infected with the indicated five different HIV-1 primary isolates and screened for the presence of HIV-1 sncRNAs in 3 contigs (contigs 2, 43 and 58) identified in HIV-1_{JR-FL} infected cultures. Cell cultures from three HIV-1 negative donors were probed for each virus and the cultures which scored positive for the respective HIV-1 sncRNAs are depicted in red squares. (b) Predicted secondary structures of sncRNA$_{LTR}$6, hybrids of sense and antisense orientated clones from contig 58, namely sncRNA$_{env}$183, sncRNA$_{env}$184 and sncRNA$_{env}$185, and the positive control siRNA-M184 $_{pol}$[29] (Supplementary Table 3) are depicted. (c) Inhibition of HIV-1 replication in primary macrophages by HIV-1 sncRNAs. Macrophages were infected with HIV-1_{JR-FL} seven days before transfection with sncRNAs. On day 0 cells were transfected with 50 nM of the indicated sncRNA (sncRNA$_{LTR}$6 (green), sncRNA$_{env}$183&185 (orange), sncRNA$_{env}$184&185 (red)) and viral replication was monitored by p24 ELISA on days 0, 5, 9 and 13 post transfection. Mock transfected (black), scrambled sncRNA (blue) and siRNA nonsense (grey) were used as negative control, and siRNA-M184 $_{pol}$[29] (open circles) as positive control.

Specific HIV-1 sncRNAs can inhibit virus replication

Considering the large number of HIV-1 sncRNAs we isolated and their variable length and orientation, it remains prudent to explore whether the various sncRNA species identified have functional properties or largely consist of degradation products which also may be generated in a sequence specific manner. Functional screening of all 67 HIV-1 specific contigs was beyond the scope of the current analysis. Here we focused in a proof-of-principle analysis on contigs 2 and 58. These contigs were chosen based on their secondary structures, which resembles precursor miRNA-like (contig 2) and siRNA-like features (contig 58). Of note, contig 58 contains sense and antisense sncRNAs. We thus chose two individual sense/antisense pairs form this contig which may form hybrids and potentially act like siRNAs (Fig. 4b). To explore if these sncRNAs have any functional impact on HIV-1 replication, primary macrophages infected with HIV-1_{JR-FL} were transfected with HIV-1 sncRNAs. While virus replication continued in mock, sncRNA and siRNA control cell cultures, HIV-1 production was potently inhibited both by the positive control siRNA-$M184_{pol}$ and the two contig 58 hybrids (Fig 4c). Both inhibited virus replication up to 90% compared to siRNA nonsense. How these RNA hybrids precisely function to downregulate HIV-1 production is currently under investigation. In contrast, the single-stranded, hairpin forming $sncRNA_{LTR}6$ had no effect on virus replication in primary macrophages in the probed setting (Fig. 4c). This preliminary analysis does not allow to define the latter as mere degradation product as we cannot rule out functional properties of this sncRNA, e.g., during earlier steps of virus replication.

Discussion

Here we report on a novel, highly efficient selection method for sncRNAs of low abundance. The latter has proven technically very challenging which may lead to an underestimation or no evidence of low abundant sncRNAs. For a model, we have chosen HIV-1 encoded sncRNAs as they were only detected in very low frequencies of 0.1-0.5% in previous studies[10,12] (and Schopman, N. *et al*., Keystone Symposia, abstract #418), if at all[13]. Our novel strategy relies on the introduction of a crucial selection step for sncRNAs homologous to HIV-1. We achieved this by adding a hybridization capture step into an improved cloning protocol for identifying sncRNAs. The hybridization capture was performed with HIV-1 ssDNA hybridization probes, covering the whole HIV-1 genome, that were attached to streptavidin beads. Applying two rounds of hybridization capture enabled us to enrich the frequencies of selected low abundant HIV-1 sncRNAs more than 100-fold over what has been reported[10,12] (and Schopman, N. *et al*., Keystone Symposia, abstract #418). Importantly, more than 70% of all obtained sncRNAs were of viral origin. This is a particular advantage of our strategy. While high-throughput sequencing techniques certainly have the capacity to overcome the limitations in identifying low abundant sncRNAs, it must still be considered that more than 99% of sequenced sncRNAs retrieved by random sequencing will not be of interest and very low abundant sncRNAs might still be missed. Our approach allows sequence specific selection with a high sensitivity. This is particularly highlighted by the fact that we succeeded in detecting antisense HIV-1 sncRNAs despite the fact that HIV-1 antisense transcripts are described to be generated only at extremely low rates[21-24].

Separate sncRNA libraries derived from infected primary cells were generated, in which 216 unique HIV-1 sncRNAs with a mean length of 43 nucleotides were identified. Although, only 8% of the clones were of lengths described for genuine miRNAs or siRNAs, it has to be considered that longer HIV-1 sncRNAs still may have regulatory functions as recently reported[25,26]. Another possibility is that those longer HIV-1 sncRNAs represent precursor molecules of miRNAs. It has been postulated that short molecules are less likely to be selected by hybridization capture[19]. However, as we show here also sncRNAs of lower length can be efficiently enriched by extracting shorter RNA molecules during size selection steps.

Amongst all the different types of sncRNAs identified in our screen, the capture of antisense HIV-1 sncRNAs was most surprising to us. Whether or not antisense HIV-1 RNAs are generated has been highly debated in the past and only few reports on HIV-1 antisense RNAs can be found in the literature[21-24]. It has been reasoned that the generation of antisense HIV-1 sncRNA might indeed be possible and occur via the HIV-1 promoter in the 3'LTR[23] or via cellular promoters downstream of the integration site[27].

Many questions regarding the generation of viral sncRNAs during the HIV-1 life cycle and their function can and need to be addressed based on our initial observations and findings. Most importantly the high number of sncRNAs identified raises the possibility that also HIV-1 RNA degradation products were selected. Indeed this cannot be ruled out entirely and functional analysis of all sncRNAs is certainly warranted. However, it is important to note that our procedure excludes the selection of degradation products generated by the classical pathways of RNA degradation which generate fragments lacking the 3'- and 5'-end modifications necessary for C-tailing and adaptor ligation[28].

Using the virus strain JR-FL we retrieved a vast number of HIV-1 sncRNAs. Of particular interest for us was to define whether these sncRNAs were specific for HIV-1_{JR-FL} only or were ubiquitously generated in HIV-1 infection. As proof-of-principle we investigated this for three contigs. Notably we found that sncRNAs of all three contigs were generated in cells infected with unrelated HIV-1 primary virus isolates, thus, confirming that the generation of these RNA species is not virus strain dependent.

Many potential functional properties of HIV-1 specific sncRNAs can be envisioned with both infection enhancing or reducing capacity. Here we report on functional assessment of sncRNA candidates from two of the 67 identified contigs. The hybridizing sense and antisense HIV-1 sncRNAs of contig 58 displayed a siRNA-like HIV-1 inhibition pattern in primary macrophages. As we demonstrate here, antisense sncRNAs appear to be generated during HIV-1 infection and thus might have the potential to downregulate HIV-1 production. This obviously raises a number of questions: Why would HIV-1 give raise to such negative regulatory RNAs? If they act

in vivo, would HIV-1 not rapidly escape and induce countermeasures? Or are these negative regulators necessary for a balanced virus production (e.g., ascertaining appropriate generation of structural proteins) or maybe in inducing latency? Now that our novel sncRNA isolation procedure provides the means to enrich and select these types of HIV-1 sncRNAs with high efficacy these functional analysis become for the first time feasible.

In summary, using hybridization capture for detection of novel sncRNAs of low abundance is a highly sensitive approach. This is particularly highlighted by the efficient enrichment of low abundant sncRNAs. More than 70% of sncRNAs we identified in our HIV-1 targeted screen were indeed derived from HIV-1 RNA demonstrating a high specificity of this enrichment by hybridization capture. The procedure proved highly reproducible and robust, as indicated by similar efficacy of sncRNA selection in all independent libraries and the identification of similar types of sncRNAs in independent experiments, respectively. Our targeted enrichment method to detect sncRNAs of low abundance by hybridization capture is highly efficient and the method would thus lend itself towards application in similar biological settings where low abundant sncRNAs need to be identified.

Supplementary Data

Supplementary Table 1: Frequencies of all sncRNA clones containing the C-tail and both 3' and 5' adaptors in each library are listed. Displayed are absolute numbers of HIV-1 sncRNAs, their percentages, and total of sequenced sncRNA clones.

			1st round of selection		2nd round of selection	
			Five HIV-1 ssDNA hybridization probes			
experiment	**cell source**	**HIV-1_{JR-FL}**	**mixed**	**separate**	**mixed**	**separate**
# 1	macrophages	+	Library A 13/101 12.9%	-	Library C 62/81 76.5%	-
# 2	$CD4^+$ T-lymphocytes	+	Library B 4/84 4.8%	-	Library D 185/227 81.5%	-
# 3	macrophages	+	not cloned	-	Library E 119/168 70.8%	-
# 4	$CD4^+$ T-lymphocytes	+	not cloned	not cloned	Library F 25/30 83.3%	Library H[(b)] 125/183 68.3%
# 5	macrophages	+	not cloned	not cloned	Library G 39/43 90.7%	Library J[(c)] 320/418 76.6%
# 6 [(a)]	macrophages	-	Library K 1/204 0.5%	-	-	-
# 7 [(a)]	$CD4^+$ T-lymphocytes	-	Library L 0/145 0.0%	-	-	-
# 8 [(a)]	macrophages	-	not cloned	-	Library M 1/185 0.5%	-
# 9 [(a)]	$CD4^+$ T-lymphocytes	-	not cloned	-	Library N 1/28 3.6%	-
# 10 [(a)]	macrophages	-	not cloned	-	Library O 8/46 17.4%	-

[(a)] for the following pairs of experiments, cells from the same donor were used: #1 and #6, #2 and #7, #3 and #8, #4 and #9, #5 and #10

[(b)] the following HIV-1 ssDNA hybridization probes were used separately: gag/pol (2), pol/env (3), sA7-LTR (5)

[(c)] all 5 HIV-1 ssDNA hybridization probes were used separately

Supplementary Table 2: SncRNA libraries generated with separate HIV-1 ssDNA hybridization probes

Library		HIV-1 ssDNA hybridization probes				
		1 TAR-gag	2 gag/pol	3 pol-env	4 env	5 sA7-LTR
H	HIV-1 sncRNAs / total sncRNA clones (%)	n.p.	29/56 (51.8%)	83/96 (86.5%)	n.p.	13/31 (41.9%)
	HIV-1 sncRNAs corresponding to the specific HIV-1 ssDNA hybridization probe	n.p.	29 (100%)	76 (91.6%)	n.p.	10 (76.9%)
	Matching HIV-1 ssDNA hybridization probe to not corresponding HIV-1 sncRNAs			1[(a)], 4, 5		1[(a)]
J	HIV-1 sncRNAs / total sncRNA clones (%)	45/67 (67.2%)	78/86 (90.7%)	67/91 (73.6%)	63/87 (72.4%)	67/87 (77.0%)
	HIV-1 sncRNAs corresponding to the specific HIV-1 ssDNA hybridization probe	43 (95.6%)	77 (98.7%)	66 (98.5%)	54 (85.7%)	64 (95.5%)
	Matching HIV-1 ssDNA hybridization probe to not corresponding HIV-1 sncRNAs	4 or 5	1[(a)]	4 or 5	1[(a)], 5	1[(a)]

n.p. = not performed

[(a)] HIV-1 sncRNAs from the RU5 genomic region were preferentially enriched as not corresponding HIV-1 sncRNAs

Supplementary Table 3: Unique HIV-1 sncRNAs

Seq No.	Library [a]	Sequence (5' - 3')	Length	Position HXB2 [b]		Polarity	Contig No.
				Start	End		
1	C	ATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTG	42	475	516	sense	1
2	H	AAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTG	33	530	562	sense	2
3	J	AAAGCTTGCCTTGTGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGA	49	530	578	sense	2
4	J	AAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACT	51	530	580	sense	2
5	J	AAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCT	53	530	582	sense	2
6	J	AAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGAT	66	530	595	sense	2
7	J	AAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACT	50	531	580	sense	2
8	D	AGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGA	47	532	578	sense	2
9	B	GCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTA	54	533	586	sense	2
10	A	CTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTG	44	534	577	sense	2
11	D	TGCCTTGAGTGCTTCAAGTAGTGTGTGCTCGTCTGTTGTGTGACTCTGGTAA	52	536	587	sense	2
12	E	AGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGT	32	543	574	sense	2
13	D	AGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAA	45	543	587	sense	2
14	A	TCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAA	39	549	587	sense	2
15	H	GTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCT	47	553	599	sense	2
16	D	GTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGAT	38	558	595	sense	2
17	C	GCCCGTCTGTTGTGTGACTCTGGTAACTAGAGAT	34	562	595	sense	2
18	H	CCCGTCTGTTGTGTGACTCTGGTAACTAGAGAT	33	563	595	sense	2
19	D	AAATCTCTAGCAGTGGCGCCCGAACAGGGACTT	33	623	655	sense	3
20	C	CCCTGTTCAGGCGCCA	16	651	636	antisense	3
21	F	AAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATT	34	779	812	sense	4
22	F	AGAGAGATGGGTGCGAGAGCGTCAGT	26	784	809	sense	4
23	G	CTTTCCCCCTGGCCTTAGCCGAATTTTTTCCCATTTATCCAATTTT	46	867	822	antisense	5
24	J	CATATAGTATGGGCAAGCAGGGAGCTAGAACGTTTCGCAGTTAATCCTGGCCTGTTAGA	59	886	944	sense	6
25	J	CATATAGTATGGGCAAGCAGGGAGCTAGGACGTTTCGCAGTTAATCCTGGCCTGTTAGAATCATCAGAAGGCTGTAT	77	886	962	sense	6
26	J	CAAGCAGGGAGCTAGAACGTTTCGCAGATAATCCTGGCCTGTTAGA	46	899	944	sense	6
27	F	TTGTGTACATCAAAGGATAGAGGTAAAAGACACCAAGGAAGCTTT	45	1047	1091	sense	7
28	C	AAAGACACCAAGGAAGCTTTAGAAAAGATAGAGGAA	36	1072	1107	sense	7
29	J	GACCTGGCTGCTGTTTCCTGTGTCAGCTGCTGCTTGCTGTGCTTTTTTCTTACTTTTGGTTTG	63	1173	1111	antisense	8
30	H	GCTTCCTCATTGATGGTCTCTCTTAGCATTTGCATGGCTGCTTGATGT	48	1407	1368	antisense	9
31	H	ATTGATGGTCTCTTTTAGCATTTGCATGGCTGCTTGATGT	40	1415	1368	antisense	9
32	G	ACAGCTTCCTCGTTGATGGTCTCTTTTAGCATTTGCATGGCTGCTTGATGT	51	1418	1368	antisense	9
33	H	AACCAAGGGGAAGTGACATAGCAGGAACTACTAGTACCCTTCAG	44	1478	1521	sense	10
34	D	GTACCCTTCAGGAACAAATAGGATGGATGACAGATAATCCACCTATCCCAGTAGGAGAAATTTATAAAAGATGGATAATCCTG	83	1511	1595	sense	10
35	J	AACAAATAGGATGGATGACAAATAATCCACCTATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGGTTT	77	1523	1599	sense	10
36	J	AATAGTAAGGATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGG	48	1605	1652	sense	11
37	H	AGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTT	38	1630	1667	sense	11
38	F	GCATTCTGGACACAAGACAAGGACCAAAGGAACCCTTTAGAGACTATGTAGACCGGTTCTT	61	1631	1691	sense	11
39	G	CATTCTGGACGTAAGACAAGGACCAAAGGAACCCTTTAGAGACTATGTAGACCGGTTCTA	60	1632	1691	sense	11
40	F	ATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGACTATGTAGACCGGTTCTA	59	1633	1691	sense	11
41	G	AGCCGAGCAAGCTTCACAGGAGGTA	25	1704	1728	sense	12
42	G	AACCTTGTTGGTCCAAAATGCGAACCCAGATCGTAAGA	38	1746	1783	sense	13
43	H	ACCTTGTTGGCCCAAAATGCGAACCCAGATTGTAAGATTA	40	1747	1786	sense	13

Seq No.	Library [a]	Sequence (5' - 3')	Length	Position HXB2 [b] Start	End	Polarity	Contig No.
44	C	ACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACTATTTTAAAAGCATTGGGACCAGCAGCCACACTAGA	74	1747	1820	sense	13
45	J	TACACTAGAAGAAATGATGACAGCATGTCAGGG	33	1812	1844	sense	13
46	H	GTCAGGGAGTGGGAGGACCCGGCCA	25	1838	1862	sense	13
47	J	AGGCAATTTTAGGAGCCAAAGAAAGATTGTTAAGTGTTTCAATTGTGG	48	1929	1976	sense	14
48	H	TTCAATTGTGGCAAAGAAGGGCACATAGCCAGAAATTGCAGGG	43	1966	2008	sense	14
49	H	TGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAAAGGAAGGA	48	2002	2049	sense	14
50	G	CACCAAATGAAAGATTGTACTGAGAGACAGG	31	2050	2080	sense	14
51	H	TCACTCTTTGGCAACGACCCCTCGTCACAATA	32	2260	2291	sense	15
52	F	GTATTAGAAGACATGGATTTGCCAGGAAGATGGAAA	36	2346	2381	sense	16
53	H	AACTCAAGACTTCTGGGAAGTTCAATTAGGAATACCACATCCCGCAGGGTT	51	2798	2848	sense	17
54	J	ACTTCTGGGAAGTTCAATTAGGAATACCACATCCCGCAGGGTTA	44	2806	2849	sense	17
55	J	AGAGAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTT	50	2854	2903	sense	18
56	J	ATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCCTTAGATAAAGACTTCAGGAAATA	69	2861	2929	sense	18
57	H	AACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCCTTAGATA	47	2867	2913	sense	18
58	G	AACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCCTTAGATAAAGACTTCAGGAAATA	63	2867	2929	sense	18
59	H	ACAGTACTGGATGTGGGTGATGCATATTTTTCAGTT	36	2868	2903	sense	18
60	H	ACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCCTTAGATA	46	2868	2913	sense	18
61	J	ACAGTACTGGGTGTGGGTGATGCATATTTTTCAGTTCCCTTAGATAAAGACCTCAGGAAATA	62	2868	2929	sense	18
62	J	ATGTGCTTCCGCAGGGATGGAAAGGATCACCAGCAA	36	2989	3024	sense	19
63	C	ATCACCAGCAATATTCCAAAGTAGCATGACA	31	3014	3044	sense	19
64	J	AACATCTGTTGAGGTGGGGGTTTACCACACCAGACAAAAAACACCAG	47	3169	3215	sense	20
65	J	ACATCTGTTGAGGTGGGGATTTACCA	26	3170	3195	sense	20
66	J	AGTCAGATTTATGCAGGGATTAAAGTAAAGCAATTATGTAAACTCCTTAGGG	52	3351	3402	sense	21
67	H	TTATGTAAACTCCTTAGGGGAACCAAAGCACTAACAGAAG	40	3384	3423	sense	21
68	H	AGAAGTAATACCGCTAACAGAAGAAGCAGAGCTAGAACTGACAG	44	3419	3462	sense	21
69	A	AACATAGTGGACAGAATACTGGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTCAATA	59	3746	3804	sense	22
70	D	CCACCTGGATTCCTGAGTGGGAGTTTGTCAATA	33	3772	3804	sense	22
71	C	CCTGGATTCCTGAGTGGGAGTTTGTCAATA	30	3775	3804	sense	22
72	J	AATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGTT	61	4334	4394	sense	23
73	J	CACATTTAGAAGGAAAAATTATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATAGA	58	4426	4483	sense	24
74	F	CCATACAGACAATGGCAGCAATTTCACCAGTACTACGGTTAAGGCCGCCTGTTGGTGGGCTGGGATT	67	4568	4634	sense	25
75	J	CGGTTAAGGCCGCCTGTTGGTGGGCTGGGATCAAGCAGGAATTTGGCATTCCCTACA	57	4603	4659	sense	25
76	J	AACATCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAATTTTA	48	4738	4785	sense	26
77	H	CCGTTGCTATTATGTCTACTATTCTTTCCCCTGCACTGTA	40	4848	4809	antisense	27
78	F	GTATGTCCGTTGCTATTATGTCTACTATTCTTTCCCCTGCACTGTA	46	4854	4809	antisense	27
79	F	AAAGGACCAGCAAAGCTTCTCTGGAAAGGTGAAGGGG	37	4935	4971	sense	28
80	J	AGATAATAGTGATATAAAAGTAGTGCCAAGAAGAAAAGTAAAGATCATTAGGGATTA	57	4985	5041	sense	29
81	G	ACTATGAAAGCACTCATCCAAGAATAAGTTCAG	33	5168	5200	sense	30
82	F	ATAAGTTCAGAAGTACACATCCCACTAGGGGATGCTAGATTAGT	44	5191	5234	sense	30
83	F	AGGAGAAAGAGACTGGCATTTGGGTCAGGGAGTCTCCATAGAATGGAGGA	50	5262	5311	sense	31
84	H	AGACTGGCATTTGGGTCAGGGAG	23	5271	5293	sense	31
85	H	GACTGGCATTTGGGTCAGGGAGT	23	5272	5294	sense	31
86	F	TGACCTAGCAGACCAACTAATTCATCTGTATTACTTTGATTGTTTTTCAG	50	5340	5389	sense	32
87	G	TTCATCTGTATTACTTTGATTGTTTTTCAGCATCTGCTAT	40	5360	5399	sense	32
88	C	ATCTGTATTACTTTGATTGTTTTTCAGCATCTGCTATAAGAAATGCCATATTAGGACATATAGTTAGTCCTAGATGTGAA	80	5363	5442	sense	32
89	H	TTGTTTTTCAGAATCTGCTATAAGAAATGCCATATTAGGACATATAGTTAGTCCTAG	57	5379	5435	sense	32

Seq No.	Library [a]	Sequence (5' - 3')	Length	Position HXB2 [b]		Polarity	Contig No.
				Start	End		
90	E	AAGAAATGCCATATTAGGACATACAGCTAGTCCTAGATGTGAATATCAAGCAGGACAT	58	5400	5457	sense	32
91	C	AAATGCCATATTAGGACATATAGTTAGTCCTAGATGTGAATATCAAGCAGGACATA	56	5403	5458	sense	32
92	D	AACTAACAGAGGATAGATGGAACAAGCCCCAGA	33	5543	5575	sense	33
93	C	ACACTAGAGCTTTTAGAGGAACTTAAAAGTGAAGCTGTTAGACATTTTCCTAGGGTCTGGCTTTT	65	5613	5677	sense	34
94	J	CATTTTCCTAGGGTCTGGCTCCATAGCTTAGGGCA	35	5655	5689	sense	34
95	H	CATTTTCCTAGGGTCTGGCTCCATAGCTTGGGGCAAT	37	5655	5691	sense	34
96	G	TACTGCAACAGCTGCTGTTTATTCATTTCAGAATTGGGTGTCAACAT	47	5746	5793	sense	35
97	H	AACAGCCGCTGTTTATTCATTTCAGAATTGGGTGTCAACAT	41	5752	5793	sense	35
98	E	GCTGCTGTTTATTCATTTCAGAATTGGGTGTCA	33	5756	5789	sense	35
99	G	AAATGGAGCCAGTAGATCCTAGCCTAG	27	5829	5855	sense	36
100	D	CTCCTATGGCAGGAAGAAGCGGAGACAGCGACGAAGAGCTCCTCAAGACA	50	5965	6014	sense	37
101	J	CTCTATCAAAGCAGTAAGTAGTGCATGTT	29	6033	6061	sense	38
102	J	CTCTATCAAAGCAGTAAGTAGTGCATGTTCA	31	6033	6063	sense	38
103	E	CAATAGTAGCATTAGTAGTAGCAATAATAATAGCAATAGTTGTGTGGTCCATAGTACTCATAGAATATAGGAAAATA	77	6084	6160	sense	39
104	H	AGTAGCATTAGTAGTAGCAATAATAATAGCAATAGTTGTGTGGT	44	6088	6131	sense	39
105	H	GTAGCATTAGTAGTAGCAATAATAATAGCAATAGTTGTGTGGTCCATA	48	6089	6136	sense	39
106	C	TAGTAGCAATAATAATAGCAATAGTTGTGTGGTCCATAGTACTCATAGAATATAGGAAAA	60	6099	6158	sense	39
107	H	AATAGCAATAGTTGTGTGGTCCATAGTACTCATAGA	36	6112	6147	sense	39
108	J	AAAATAGATAGGTTAATTGATAGAATAAGAGAGAGT	36	6173	6208	sense	40
109	H	ATTATCAGCACTTGTGGAGAGGGGGCACCTTGCTCCTTGGAAT	43	6241	6295	sense	41
110	H	AATCAGCACTTGTGGAGAGGGGGCACCTTGCTCCTTGGAA	40	6243	6294	sense	41
111	H	CTCCTTGGGATATTAATGATCTGTAGTGCTGTAGAA	36	6285	6320	sense	41
112	D	TGGAATAATAATGATCTGTAGTGCTGTAGAAAAGTTGTGGGTCACAGTCTAT	52	6290	6341	sense	41
113	C	ATATTAATGATCTGTAGTGCTGTAGAAAAGTTGTGGGTCACAGT	44	6294	6337	sense	41
114	J	ATATTAATGATCTGTAGTGCTGTAGAAAAGTTGTGGGTCACAGT	44	6294	6337	sense	41
115	H	AATAATGATCTGTAGTGCTGGAGAAAAGTTGTGGGTCACAGTCT	44	6296	6339	sense	41
116	D	TAATGATCTGTAGTGCTGTAGAAAAGTTGTGGGTCACAGTCT	42	6298	6339	sense	41
117	G	AATGATCTGTAGTGCTGTAGAAAAGTTGTGGG	32	6299	6330	sense	41
118	G	ATCTGTAGTGCTGTAGAAAAGTTGTGGGT	29	6303	6331	sense	41
119	J	CAGTCTATTGTGGGGTACCTGTGTGGAGAGAAGCAACCACCACTCTATTTTGTGCATCA	59	6334	6392	sense	41
120	C	GCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGA	56	6402	6457	sense	42
121	H	TATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGA	53	6405	6457	sense	42
122	E	ACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGA	39	6419	6457	sense	42
123	E	ACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCTAACCCACAAGAAGTAGTATTGGAAAATGT	72	6419	6490	sense	42
124	H	AATGTTTGGGCCACACATGCCTGTATACCCACAGA	35	6423	6457	sense	42
125	D	ATGTTCGGGCCACACATGCCTGTGTACCCACAGA	34	6424	6457	sense	42
126	H	TTTGGGCCACACATGCCTGTGTACCCACAGA	31	6427	6457	sense	42
127	H	TTGGGCCACACATGCCTGTGTACCCACAGA	30	6428	6457	sense	42
128	D	CCACACATGCCTGTGTACCCACAGA	25	6433	6457	sense	42
129	E	CCACACATGCCTGTGTACCCACAGA	25	6433	6457	sense	42
130	A	ACACATGCCTGTGTACCCACAGA	23	6435	6457	sense	42
131	E	CACATGCCTGTGTACCCACAGA	22	6436	6457	sense	42
132	H	CACATGCCTGTGTACCCACAGA	22	6436	6457	sense	42
133	G	ACATGCCTGTGTACCCACAGA	21	6437	6457	sense	42
134	F	TAACATGTGGAAAAATAACATGGTAGAACAGATGCAGGAGGAT	43	6503	6545	sense	43
135	D	AACAGATACAGGAGGATATAATCAGTTTATGGGATCAAAGCCTAA	45	6529	6573	sense	43

Seq No.	Library [a]	Sequence (5' - 3')	Length	Position HXB2 [b] Start	End	Polarity	Contig No.
136	H	TAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAA	45	6547	6591	sense	43
137	G	AATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTTACCCCACTCTGTGTTACTTTA	64	6548	6613	sense	43
138	D	GTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCCCACTCTGTGTTA	54	6553	6606	sense	43
139	J	ACCCCACTCTGTGTTACTTTAAATTGCAAGGATGTG	36	6591	6626	sense	43
140	F	CTCTGTGTTACTTTAAATTGCAAGGATGTGAATGCTACTAATA	43	6597	6642	sense	43
141	J	CTTTAAATTGCAAGGATGTGAATGCTACT	29	6607	6635	sense	43
142	H	TTGCAATGATGTGAATGCTACTAATACCACTAGTGGTAGCGAGG	44	6614	6660	sense	43
143	J	GTGGTGATATTGAAAGAGCAGTTTTTTATTTCT	33	6712	6680	antisense	44
144	J	AACTTGATGTAGTACCAATAGGATAATAATAATACCAGCTATAGGTTGATA	51	6757	6807	sense	45
145	A	CCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAGTGTAATG	49	6864	6912	sense	46
146	E	CATTATTGTGCCCCGGCTGGTTTTGCGATT	30	6870	6899	sense	46
147	D	CATTATTGTGCCCCGGCTGGTTTTACGATTCTAAAGTG	38	6870	6907	sense	46
148	C	CATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAGTGTAATG	43	6870	6912	sense	46
149	A	CATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAGTGTAATGATAAGACGTT	53	6870	6922	sense	46
150	A	CCCCGGCTGGTTTTGCGATTCTAAAGTGTAAT	32	6880	6911	sense	46
151	C	TTCTAAAGTGTAATGATAAGACCTTCAATGGAAAAGGACCATGTATTT	48	6898	6945	sense	46
152	J	TCTAAAGTGTAATGATAAGACGTTCAATGGAAAAGG	36	6899	6934	sense	46
153	C	ACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGCTAAATGGCAGTCTAGCAGA	66	6959	7024	sense	47
154	G	AATTAGATCTGACAATTTCACGAACAATGCTAAAACCA	38	7037	7074	sense	48
155	C	ATTAGATCTGACAATTTCACGAACAATGCTAA	32	7038	7069	sense	48
156	E	AATTGTAATGCACAGTTTTAATTGTGGAGGAGAATTTTTCTACTGT	46	7334	7379	sense	49
157	J	AACTGTTTAATAGTACTTGGAATAATAATACTGAAGGGTCAAATT	45	7390	7443	sense	50
158	B	ATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTATG	37	7494	7530	sense	51
159	H	CATGTGGCAGGAAGTAGGAAAAGCAATGTATG	32	7499	7530	sense	51
160	D	ATGTGGCAGGAAGTAGGAAAAGCAATGTATG	31	7500	7530	sense	51
161	J	ATGTGGCAGGAAGTAGGAAAAGCGATGTATG	31	7500	7530	sense	51
162	D	GACCTGGAGGAGGAGATATGAGGGACAACTGGAGAAGTGA	40	7630	7669	sense	52
163	C	ACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTG	38	7631	7668	sense	52
164	H	ACATCACTTCTCCAATTGCCCCTCATATCT	30	7672	7643	antisense	52
165	E	CTATTATTGTCTGGTATAGTGCAACAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCGTATGTTGCAACTCACAGTCTGGGG	89	7851	7939	sense	53
166	D	CTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTG	36	8018	8053	sense	54
167	D	AACTCATTTGCACCACTGCTGTGCCTTGGAATGCTA	36	8026	8061	sense	54
168	H	CTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGATA	45	8041	8085	sense	54
169	J	GATTTGGAATAACATGACCTGGATGGAGTGGGT	33	8087	8119	sense	55
170	E	AACATGACCTGGATGGAGTAGAAAAGAGAAATTGACAATTACACAAGCGAA	51	8097	8147	sense	55
171	J	TGACCTGGATGGAGTGGGAAAGAGAAATTTACAATTACACA	41	8101	8141	sense	55
172	J	AACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTGACATA	56	8194	8249	sense	56
173	J	ATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTGACATA	46	8204	8249	sense	56
174	J	AGTTTGTGGAATTGGTTTGACATAACAAAATGGCTG	36	8226	8261	sense	56
175	E	AGGCTTGATAGGTTTAAGAATAGTTTTTACTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATTT	67	8294	8360	sense	57
176	C	TAGTTTTTACTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATACTCACCATTATCGTTTCAGA	66	8314	8379	sense	57
177	J	ATACTCACCATTATCGTTTCAGA	23	8357	8379	sense	57
178	H	CTCACCATTATCGTTTCAGA	20	8360	8379	sense	57
179	J	TCACCATTATCGTTTCAGA	19	8361	8379	sense	57
180	J	CCGTTCACTAATCGTTCGGATCTGTCTCTGTCTCTCTCT	39	8476	8438	antisense	58
181	E	TAATCGTCCGGATCTGTCTCTGTCTCTCT	29	8468	8440	antisense	58

Seq No.	Library [a]	Sequence (5' - 3')	Length	Position HXB2 [b]		Polarity	Contig No.
				Start	End		
182	J	TTCACTAATCGTCCGGATCTGTCTCTGTCTCTCT	34	8473	8440	antisense	58
183	E	CGTTCACTAATCGTCCGGATCTGTCTCTGTCTCTCT	36	8475	8440	antisense	58
184	J	TTCACTAATCGTCCGGATCTGTCTCTGTCTCT	32	8473	8442	antisense	58
185	H	CAGATCCGGACGATTAGTGAACGGATT	27	8453	8479	sense	58
186	E	CACTTATCTGGGTCGACCTGCGGAGCCTGTG	31	8485	8515	sense	59
187	H	ACTTATCTGGGTCGACCTGCGGAGCCTGTG	30	8486	8515	sense	59
188	J	ACTTATCTGGGTCGACCTACGGAGCCTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGACTGTAACGAG	78	8486	8563	sense	59
189	C	CTCTTGACTGTAACGAGGATTGTGGAA	27	8547	8573	sense	59
190	H	GATTGTGGAACTTCTGGGACGCAGGGAGTGAGAAGTCCTGAAATATCGGTGGG	53	8564	8616	sense	59
191	G	CTTGCTCAATGCCACAGCCATA	22	8663	8684	sense	60
192	J	CTTGCTCAATGCCACAGCCATAG	23	8663	8685	sense	60
193	E	CAAAACGTAGTGTGCCTGGATGGTCTACTGTAAGGGA	37	8813	8849	sense	61
194	G	AAAACGTAGTGTGCCTGGATGGTCTACTGTAAGGGAAAG	39	8814	8852	sense	61
195	G	CAGGTCTCGAGATACTGCT	19	8907	8889	antisense	62
196	E	TTGCTCCGTGTTTTTCCAGGTCTCGAGATACTGCT	35	8923	8889	antisense	62
197	E	GATTGCTCCATGTTTTTCCAGGTCTCGAGATACTGCT	37	8925	8889	antisense	62
198	C	CGAGACCTGGAAAAACATGGAGCAATCACAAGTAGCAATACAGCAGCTACCAATGCTGATTGTGCCTGACTAGAAGCATA	80	8899	8978	sense	62
199	B	AAGCACAAGAGGATGAGGAGGTGGGTTTTCCAGTCAGACCTCAGGTA	47	8972	9018	sense	62
200	C	CCAATGACTTACAAGGGAGCTGTAGATCTTAGCCACTTTTTAAAAGA	47	9028	9074	sense	63
201	F	CACACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGT	51	9144	9194	sense	64
202	G	TGATTGGCAGAACTACACACCAGGGCCAGGAATCAGATTTCCATTGACCTTTGGATGGTGCT	62	9162	9223	sense	64
203	E	CTACACACCAGGGCCAGGAATCAGATTTCCATTGACCTTTGGATGGTGCTTCAAGCTAG	59	9174	9232	sense	64
204	D	GATTTCCACTGACCTTTGGATGGTGCTTCAAGCTAGTACCAGTTGAGCCA	50	9197	9246	sense	64
205	J	CATTGACCTTTGGATGGTGCTTCAAGCTAGTACCAGTTGAGCCA	44	9203	9246	sense	64
206	D	CTGACCTTTGGATGGTGCTTCAAGCTAGTACCAGTTGAG	39	9205	9243	sense	64
207	D	CTGACCTTTGGATGGTGCTTCAAGCTAGTACCAGTTGAGCCA	42	9205	9246	sense	64
208	G	AACAAGCAGTTGTTCTCTCCTT	22	9290	9269	antisense	65
209	D	GCTTGTTACACCCTATGAGCCAGCATGGGATGGGCGA	37	9284	9320	sense	65
210	D	GCTTGTTACACCCTATGAGCCAGCATGGGATGGACGACCCGGAGAGGGAA	50	9284	9333	sense	65
211	E	ATGGGATAGAGGACCCGGAGAAGGAAGTGTTAGAGTGGAGGTTTGACAG	49	9308	9356	sense	65
212	H	TCATCACGTGGCCCGAGAGCTGCATCCGGAGTACT	35	9369	9403	sense	66
213	C	TCACGCGGCCCGAGAGCTGCATCCGGAGTACTACAAGGA	39	9372	9410	sense	66
214	D	TCACGTGGCCCGAGAGCTGCATCCGGAGTACTACAAGGA	39	9372	9410	sense	66
215	C	CGCCTCCCTGGAAAGT	16	9467	9452	antisense	67
216	H	CCTCCCGCGCAGGCCACG	18	9483	9466	antisense	67

[a] libraries A, C, E, G and J are derived from macrophages and libraries B, D, F and G from $CD4^+$ T-lymphocytes

[b] positions of sncRNAs are based on the reference HIV-1 strain HXB2 (GenBank accession number K03455)

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This study was funded by the Swiss National Science Foundation (grant no. 31003A_127317 to MF and HFG, grant no 324730-130865 to HFG) and by the Hartmann Müller-Stiftung to MF.

Authors' contributions

CFA carried out the analysis of the sequence data, performed the alignments and the statistical analyses and wrote the manuscript. VV established the PCR for the sncRNAs, performed the transfection and HIV-1 inhibition experiments. BN prepared the libraries, cloned and sequenced sncRNAs and performed the HIV-1 sncRNA-specific PCRs. BJ assisted in the evaluation of the sequences. FG assisted in the analysis and illustration of the data. PR provided the full length sequences of env of the primary virus isolates. JP assisted in sncRNA transfection experiments. AT participated in the design of the study and writing of the manuscript. HFG participated in the design of the study and in the analysis. KJM participated in the analysis, design of the study and writing of the manuscript. MF conceived the study, participated in its design and coordination and in the analysis. This article is dedicated to the memory of Marek Fischer, who died in December 2010. All other authors contributed to the writing of the manuscript, read and approved the manuscript.

Methods

Viruses

Primary HIV-1 isolates[20] were derived from patients' peripheral blood mononuclear cells (PBMC) by co-culturing patient $CD4^+$ T-lymphocytes with stimulated, CD8-depleted PBMC as previously described[30]. Patients were enrolled in the Zurich primary HIV infection (ZPHI) study (http://clinicaltrials.gov: NCT00537966) and written informed consent was obtained from all participants. Viral replication was assessed from culture supernatants by p24 ELISA (adapted from[31]). $TCID_{50}$ of primary isolates and PBMC grown HIV-1_{JR-FL} virus stocks was estimated as described[32].

Cells and infection

PBMC from three healthy donors were isolated, $CD8^+$ T-cell depleted and cultured as described previously[32]. For stimulation of $CD4^+$ T-lymphocytes, cells were divided in 3 parts and cultured in the presence 0.5 µg/ml phytohemagglutinin (PHA), 5 µg/ml PHA, or anti-CD3 antibody OKT3 hybridoma supernatant[32]. After 72 h, cells from all 3 stimulations were combined. Cells were infected with HIV-1_{JR-FL} (MOI = 0.01), harvested 7 days post infection and lyzed using QIAzol lysis reagent (Qiagen). Virus replication was monitored by p24 ELISA.

For the generation of macrophages, primary human monocytes were isolated from $CD8^+$ T-cell depleted PBMC using positive selection with anti-CD14-coated magnetic beads (Miltenyi Biotech). Monocytes ($3x10^6$ cells per T25 flask) matured to macrophages in the presence of 0.02 µg/ml human M-CSF (macrophage colony stimulating factor, PeproTech). Macrophages were maintained in RPMI-1640 supplemented with 10% FCS, 1% penicillin/streptomycin, 5% MCM (macrophage conditioned medium, sterile-filtered), 5% human serum and 0.02 µg/ml M-CSF (the latter three ingredients were added only during the first 6-8 days). After 14 days of maturation, macrophages were infected with HIV-1_{JR-FL} (MOI = 0.01). Medium was changed 16 h post infection and virus replication was monitored by p24 ELISA. After another 14 days, cells were harvested and lyzed using QIAzol lysis reagent (Qiagen).

Isolation of the low molecular weight RNA fraction

Lyzed cells were homogenized with QIAshredder (Qiagen) and the extraction of small RNA (<200 nt) was performed using miRNeasy Mini Kit (Qiagen). Briefly, the flow through after homogenization was supplemented with 140 µl chloroform, thoroughly mixed and centrifuged for 15 min at 13,500 g. The upper phase containing the RNA was supplemented with 1 volume ethanol (70%) and transferred to a spin column. After centrifugation for 15 s at 8,000 g, the column containing RNA fragments >200 nt was discarded, while the flow through with the small RNA fragments (<200 nt) was saved. 450 µl of ethanol (100%) were added to the flow through, mixed thoroughly and transferred to an RNeasy MinElute spin column (Qiagen). After centrifugation for 15 s at 8,000 g, the column was washed with buffer RPE and 80% ethanol according to the manufacturers' instructions. RNA was eluted in 40 µl RNase-free water. The RNA was quantified with Nanodrop 1000 (Thermo Scientific).

Adaptor addition and cDNA synthesis

An aliquot (15 µl) of the low molecular weight fraction of extracted RNA was C-tailed for 15 min at 37°C using 7.5 units E.*coli* Poly(A) Polymerase (New England Biolabs) and 0.75 mM CTP (Connectorate). The synthesis of C-tails was blocked by addition of 0.5 mM Cordycepin (5'-triphoshpate sodium salt, Sigma-Aldrich) and 2.5 units E.*coli* Poly(A) Polymerase, and incubation for 15 min at 37°C. At the same time, C-tailed RNA was treated with 15 U DNase (DNase I recombinant, RNase-free; Roche). Afterwards, precipitation was performed by adding 1 volume isopropanol, 0.2 M sodium acetate and 4 µl precipitation carrier Dr. Gentle (Takara) and centrifuged for 30 min at 16°C and 16,000 g. The pellet was washed with 80% ethanol and eluted in 20 µl H_2O. Subsequently, the 5'-end was ligated to an 2' O-methylated RNA adaptor 5'-AUCGGAACAUCCAGACAUAACA-3' using 40 U T4 RNA ligase (New England Biolabs), 4 µM adaptor RNA and 60 U RNaseOut (Intvitrogen) (15 min at 37°C and over night at 16°C). This was followed by precipitation as described above and elution in 10 µl H_2O. cDNA was generated using M-MuLV Reverse Transcriptase (Finnzymes) and the 3' linker primer mf331 5'-ACCAGAGTGCGAGTAGGAAGATTGGGGGGGGG-3' partly complementary to the C-tail of the RNA in a final volume of 20 µl. Briefly, RNA and 5 µM primer were denaturated for 5 min at 95°C followed by incubation on ice for at least 2 min. The

enzyme-buffer-dNTP (400 U M-MuLV Reverse Transcriptase, 2.5 mM dNTP) mixture was added and the reaction was incubated for 60 min at 37°C. Amplification of 2 µl cDNA was executed with JumpStart Taq ReadyMix (Sigma-Aldrich) for 15 cycles using 1 µM 5' adaptor primer mf311 5'-ATCGGAACATCCAGACATAACA-3' and 1.5 mM $MgCl_2$ in a final volume of 50 µl (95°C-5'; 15x (95°C-5"; 52°C-5"; 72°C-40")).

A second round of PCR with 25 cycles was performed using 1 µl of a 1:10 dilution of the first PCR product. Again JumpStart Taq ReadyMix (Sigma-Aldrich) supplemented with 1.5 mM $MgCl_2$ and 1 µM of each 5' and 3' adaptor primers mf311 and mf315 5'-ACCAGAGTGCGAGTAGGAAGATT-3' was used in a final volume of 30 µl (95°C-2'; 25x (95°C-5"; 52°C-5"; 72°C-20")). Amplicons were precipitated with isopropanol and dissolved in $TENT_{5/200}$ (10 mM Tris-HCl, 5 mM EDTA, 0.2 M NaCl, 0.1% Triton) in a final volume of 40 µl. All RNA and DNA oligonucleotides were synthesized by Microsynth.

Generation of HIV-1 DNA/Streptavidin beads for selection of HIV-1 sncRNAs

The HIV-1_{JR-FL} plasmid[33] was used as template and amplified with HIV-1 specific biotinylated primers, using the HotStartTaq Master Mix Kit (Qiagen) supplemented with 1.5 mM $MgCl_2$ in a final volume of 50 µl (95°C 15'; 40x (95°C 10" - 57°C 10" - 72°C 4'); 72°C 7'). Five amplicons were generated using the following primers that are biotinylated at the 5'-end: 1) TAR to gag (position according to HIV-1_{HXB2} (GenBank accession number K03455): 455-1,658) with the primers mf271 5'-GGTCTCTCTGGTTAGACCAGATTTGA-3' and sk39 5'-TTTGGTCCTTGTCTTATGTCCAGAATGC-3', 2) gag to pol (1,273-4,837) with mf219 5'-AAGGCTTTCAGCCCAGAAGTAATACCCATGTT-3' and mf255 5'-ATGTCTACTATTCTTTCCCCTGCA-3', 3) pol to env (4,758-6,741) with mf254 5'-CAAATGGCAGTATTCATCCACAA-3' and mf237 5'-ATTCTTCCTGATCCCCTTCACTCTCAT-3', 4) env (5,956-8,421) with mf1 5'-CTTAGGCATCTCCTATGGCAGGAA-3' and mf2 5'-TTCCTTCGGGCCTGTCGGGTCCC-3' and 5) splice acceptor 7 (sA7) to the 3'LTR (8,317-9,700) with mf214 5'-TTTTTGCTGTACTTTCTATAGTGAATAGAGTTA-3' and cr2 5'-TGACTAAAAGGGTCTGAGGGATCTCTAGTTACCAG-3'. All primers were used in a final concentration of 1 µM. The PCR products were purified (Qiaquick,

Qiagen) and eluted in 10 mM Tris-HCl (pH 8.5). All primers were synthesized by Microsynth.

Biotinylated DNA was attached to streptavidin beads (Roche). Either 400 ng of biotinylated DNA from each PCR was used separately, or in combination (5 x 400 ng) for preparation of the beads. Briefly, 25 µg beads were washed with $TENT_{100}$ buffer (10 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl, pH 7.5, 0.1% Triton) and resuspended in 75 µl 2x$TENT_{100}$. Denaturated amplicons (5', 95°C) were added to the beads and the volume was adjusted to 150 µl with H_2O. DNA was immobilized by 30 min incubation with the beads at 37°C. Streptavidin-biotinylated, single-stranded DNA complexes were achieved by heating to 90°C for 1 min. The attachement-dehybridization procedure was repeated once. Streptavidin-biotinylated-ssDNA complexes were washed 3 times with $TENT_{1000}$ (10 mM Tris-HCl, 1 mM EDTA, 1 M NaCl, pH 7.5, 0.1% Triton) and 3 times with $TENT_{100}$. They were stored in $TENT_{100}$ at 4°C.

Selection of HIV-1 sncRNAs

For the hybridization of amplified HIV-1 sncRNAs to the Streptavidin-biotinylated-ssDNA complexes, 10 µl of these beads were added to the amplified HIV-1 sncRNAs and incubated for 3 min at 95°C followed by a cool down to 50°C over night on a head to tail wheel. Beads were washed 4 times with pre-warmed (50°C) $TENT_{5/200}$ buffer. Annealed amplified HIV-1 sncRNAs were eluted from the beads by adding 15 µl Tris-HCl buffer (10 mM Tris-HCl, pH 8.5) and heating for 5 min at 95°C. Beads and eluted sncRNA were separated by magnetic separation. HIV-1 sncRNAs were amplified using JumpStart Taq ReadyMix (Sigma) supplemented with 1.5 mM $MgCl_2$ and 1 uM of each adaptor-specific primers mf311 and mf315 (95°C-2'; 30x (95°C-5"; 52°C-5"; 72°C-30")) in a final volume of 30 µl. Amplicons were size-separated using a 3% MetaPhor agarose gel. DNA with a length of 50-110 bp was extracted from gel using GenElute Agarose Spin Columns (Sigma) according to the manufacturer's instruction. When two selection steps were performed, eluate was precipitated with isopropanol and the hybridization and size selection steps were repeated. Eluates were precipitated with isopropanol and eluted in 15 ul H_2O.

Cloning and Sequencing of HIV-1 sncRNAs

Amplified and selected HIV-1 sncRNAs were ligated into the vector pDrive using the QIAGEN PCR Cloning kit (Qiagen) according to manufacturer's instruction. Single clones were sequenced in one direction with the primer T7 using BigDye chain terminator chemistry and the automated sequencer ABI 3100 (Applied Biosystems). Sequences were controlled for the presence of both adaptor sequences, which were subsequently deleted to obtain the sncRNA sequence. This analysis was performed using the software BioEdit[34]. All sncRNA sequences were aligned to the reference strains HIV-1_{HXB2} and HIV-1_{JR-FL} using the software DNAstar (DNA-star Madison). Sequences with >90% homology to the reference strain HIV-1_{JR-FL} were considered HIV-1 specific. FASTA[35] was chosen for further nucleotide similarity searches. Classification of small RNA sequences was based on sequence analysis using the GenBank database (http://www.ncbi.nlm.nih.gov/genbank/), the miRNA registry database (http://www.mirbase.org/) and the human tRNA database (http://gtrnadb.ucsc.edu/). Secondary structures of selected HIV-1 sncRNA were predicted with RNAstructure 5.2[36]. SncRNA sequences smaller than 16 nucleotides were not included in our analysis.

Statistical analyses

Statistical analyses were performed using GraphPad Prism5.0 software. The two-tailed Chi square test and the Wilcoxon rank sum test were used for binary and cardinal data, respectively. $p<0.05$ was considered statistically significant.

Transfection of primary macrophages with HIV-1 sncRNAs

Maturated macrophages were generated and infected with HIV-1_{JR-FL} as described above. Seven days after infection cells were transfected with HIV-1 sncRNAs using jetPRIME transfection reagent (Polyplus-Transfection). Briefly, medium was replaced by Opti-MEM® I Reduced Serum Media (Invitrogen) and the transfection mix was added to the cells according to the manufacturer's instructions. After four hours, 10 % FCS (Invitrogen) was added. The next day the transfection medium was replaced by RPMI-1640 supplemented with 10% FCS and 1% penicillin/streptomycin.. Virus replication was observed by p24 ELISA. Following oligonucleotides were used for sncRNA transfection: sncRNA$_{LTR}$6; sncRNA$_{env}$183; sncRNA$_{env}$184; sncRNA$_{env}$185 (Supplementary Table 3). Control siRNA labelled with AlexaFluor488 (AllStars

negative controls, Qiagen), here named as siRNA nonsense, was used as control for the transfection efficiency and negative control for virus inhibition, whereas siRNA-M184$_{pol}$ was chosen as positive control as previously described[29].

Detection of HIV-1 sncRNAs in cells infected with primary HIV-1 isolates

HIV-1 sncRNAs detected in most or all of the libraries, were screened in terms of their presence in primary cells infected with primary HIV-1 isolates. Infection of CD8-depleted PBMC, RNA isolation, C-tailing and reverse transcription was performed as described above. HIV-1 sncRNAs were amplified. Briefly, PCR reactions contained 1x PCR buffer (Sigma), 0.67x Eva-Green (Brunschwig), 3 mM $MgCl_2$, 0.2 mM dNTPs, 1 uM of primer mf315, 1 uM of respective HIV-1 sncRNA-specific primer (mf382 5'- ATAAAGCTTGCCTTGAGTG-3', mf178 5'-ATGGTAGAACAGATGCATGAGGATATAAT-3', CA-179 5'-CGTTCACTAATCGTCCGGATCTGTC-3') and 0.03 U/µl JumpStart Taq DNA Polymerase (Sigma). PCR was performed as follows: 95°C-2'; 50x (94°C-10"; 55°C-10"; 72°C-40"). Amplicons were loaded on 3% MetaPhor agarose gel and separated by electrophoresis mobility.

References

1. Fire, A. et al., Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391, 806 (1998).
2. Farazi, T. A., Juranek, S. A., and Tuschl, T., The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. Development 135, 1201 (2008).
3. Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N., Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet 9, 102 (2008).
4. Huttenhofer, A. and Vogel, J., Experimental approaches to identify non-coding RNAs. Nucleic Acids Res 34, 635 (2006).
5. Lu, C. et al., Elucidation of the small RNA component of the transcriptome. Science 309, 1567 (2005).
6. Girard, A., Sachidanandam, R., Hannon, G. J., and Carmell, M. A., A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 442, 199 (2006).
7. Kong, W., Zhao, J. J., He, L., and Cheng, J. Q., Strategies for profiling microRNA expression. J Cell Physiol 218, 22 (2009).
8. Pfeffer, S. et al., Identification of virus-encoded microRNAs. Science 304, 734 (2004).
9. Mrazek, J., Kreutmayer, S. B., Grasser, F. A., Polacek, N., and Huttenhofer, A., Subtractive hybridization identifies novel differentially expressed ncRNA species in EBV-infected human B cells. Nucleic Acids Res 35, e73 (2007).
10. Pfeffer, S. et al., Identification of microRNAs of the herpesvirus family. Nat Methods 2, 269 (2005).
11. Parameswaran, P. et al., Six RNA viruses and forty-one hosts: viral small RNAs and modulation of small RNA repertoires in vertebrate and invertebrate systems. PLoS Pathog 6, e1000764 (2010).
12. Yeung, M. L. et al., Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid. Nucleic Acids Res 37, 6575 (2009).
13. Lin, J. and Cullen, B. R., Analysis of the interaction of primate retroviruses with the human RNA interference machinery. J Virol 81, 12218 (2007).
14. Klase, Z. et al., HIV-1 TAR element is processed by Dicer to yield a viral micro-RNA involved in chromatin remodeling of the viral LTR. BMC Mol Biol 8, 63 (2007).
15. Ouellet, D. L. et al., Identification of functional microRNAs released through asymmetrical processing of HIV-1 TAR element. Nucleic Acids Res 36, 2353 (2008).
16. Bennasser, Y., Le, S. Y., Benkirane, M., and Jeang, K. T., Evidence that HIV-1 encodes an siRNA and a suppressor of RNA silencing. Immunity 22, 607 (2005).
17. Omoto, S. et al., HIV-1 nef suppression by virally encoded microRNA. Retrovirology 1, 44 (2004).
18. Kaul, D., Khanna, A., and Suman, Evidence and nature of a novel miRNA encoded by HIV-1. Proc Indian Natl Sci Acad 72, 91 (2006).
19. Mamanova, L. et al., Target-enrichment strategies for next-generation sequencing. Nat Methods 7, 111 (2010).
20. Rieder, P. et al., HIV-1 transmission after cessation of early antiretroviral therapy among men having sex with men. AIDS 24, 1177 (2010).

21. Michael, N. L. et al., Negative-strand RNA transcripts are produced in human immunodeficiency virus type 1-infected cells and patients by a novel promoter downregulated by Tat. J Virol 68, 979 (1994).
22. Landry, S. et al., Detection, characterization and regulation of antisense transcripts in HIV-1. Retrovirology 4, 71 (2007).
23. Ludwig, L. B. et al., Human Immunodeficiency Virus-Type 1 LTR DNA contains an intrinsic gene producing antisense RNA and protein products. Retrovirology 3, 80 (2006).
24. Lefebvre, G. et al., Analysis of HIV-1 Expression Level and Sense of Transcription by High-Throughput Sequencing of the Infected Cell. J Virol 85, 6205 (2011).
25. Kapranov, P. et al., RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316, 1484 (2007).
26. Wang, X. et al., Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature 454, 126 (2008).
27. Lenasi, T., Contreras, X., and Peterlin, B. M., Transcriptional interference antagonizes proviral gene expression to promote HIV latency. Cell Host Microbe 4, 123 (2008).
28. Hafner, M. et al., Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. Methods 44, 3 (2008).
29. Huelsmann, P. M., Rauch, P., Allers, K., John, M. J., and Metzner, K. J., Inhibition of drug-resistant HIV-1 by RNA interference. Antiviral Res 69, 1 (2006).
30. Trkola, A. et al., Delay of HIV-1 rebound after cessation of antiretroviral therapy through passive transfer of human neutralizing antibodies. Nat Med 11, 615 (2005).
31. Moore, J. P., McKeating, J. A., Weiss, R. A., and Sattentau, Q. J., Dissociation of gp120 from HIV-1 virions induced by soluble CD4. Science 250, 1139 (1990).
32. Rusert, P. et al., Virus isolates during acute and chronic human immunodeficiency virus type 1 infection show distinct patterns of sensitivity to entry inhibitors. J Virol 79, 8454 (2005).
33. Koyanagi, Y. et al., Dual infection of the central nervous system by AIDS viruses with distinct cellular tropisms. Science 236, 819 (1987).
34. Hall, T. A., Bioedit: A user-friendly biological sequence alignment editor and analysis program for Windows 95=98 NT. Nucleic Acids Symp Ser 41, 95 (1999).
35. Pearson, W. R. and Lipman, D. J., Improved tools for biological sequence comparison. Proc Natl Acad Sci U S A 85, 2444 (1988).
36. Reuter, J. S. and Mathews, D. H., RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics 11, 129 (2010).

3. Monitoring HIV-1 RNA transcription patterns during cART in acutely infected patients to assess the effect of early treatment on cellular viral reservoirs

Published as Schmid *et al*., PLoS One. 2010 Oct 12;5(10)

OPEN ACCESS Freely available online

PLoS one

Profound Depletion of HIV-1 Transcription in Patients Initiating Antiretroviral Therapy during Acute Infection

Adrian Schmid[•], Sara Gianella[•], Viktor von Wyl, Karin J. Metzner, Alexandra U. Scherrer, Barbara Niederöst, Claudia F. Althaus, Philip Rieder, Christina Grube, Beda Joos, Rainer Weber, Marek Fischer*[¶], Huldrych F. Günthard*[¶]

Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zürich, University of Zürich, Zürich, Switzerland

Abstract

Background: Although combination antiretroviral therapy (cART) initiated in the acute phase of HIV-1 infection may prevent expansion of the latent reservoir, its benefits remain controversial. In the current study, HIV-1 RNA transcription patterns in peripheral blood mononuclear cells (PBMC) were monitored during acute cART to assess the effect of early treatment on cellular viral reservoirs.

Methodology/Principal Findings: Acutely HIV-1 infected patients (n = 24) were treated within 3–15 weeks after infection. Patients elected to cease treatment after ≥1 year of therapy. HIV-1 DNA (vDNA), HIV-1 RNA species expressed both in latently and productively infected cells, unspliced (UsRNA), multiply spliced (MsRNA-tatrev; MsRNA-nef), and PBMC-associated extracellular virion RNA (vRex), expressed specifically by productively infected cells, were quantified in PBMC by patient matched real-time PCR prior, during and post cART. In a matched control-group of patients on successful cART started during chronic infection (n = 15), UsRNA in PBMC and vDNA were measured cross-sectionally. In contrast to previous reports, PBMC-associated HIV-1 RNAs declined to predominantly undetectable levels on cART. After cART cessation, UsRNA, vRex, and MsRNA-tatrev rebounded to levels not significantly different to those at baseline ($p>0.1$). In contrast, MsRNA-nef remained significantly lower as compared to pretreatment ($p = 0.015$). UsRNA expressed at the highest levels of all viral RNAs, was detectable on cART in 42% of patients with cART initiated during acute infection as opposed to 87% of patients on cART initiated during chronic infection (Fisher's exact test; $p = 0.008$). Accordingly, UsRNA levels were 105–fold lower in the acute as compared to the chronic group.

Conclusion: Early intervention resulted in profound depletion of PBMC expressing HIV-1 RNA. This is contrary to chronically infected patients who predominantly showed continuous UsRNA expression on cART. Thus, antiretroviral treatment initiated during the acute phase of infection prevented establishment or expansion of long-lived transcriptionally active viral cellular reservoirs in peripheral blood.

Citation: Schmid A, Gianella S, von Wyl V, Metzner KJ, Scherrer AU, et al. (2010) Profound Depletion of HIV-1 Transcription in Patients Initiating Antiretroviral Therapy during Acute Infection. PLoS ONE 5(10): e13310. doi:10.1371/journal.pone.0013310

Editor: Lishomwa C. Ndhlovu, University of California San Francisco, United States of America

Received June 22, 2010; **Accepted** September 10, 2010; **Published** October 12, 2010

Funding: This work was supported by the Swiss National Science Foundation (grant # 324730-116035 and -130865 to HFG, grant # 3100A0-112670 to MF and HFG), by grants from the Stiftung für Wissenschaftliche Forschung an der Universität Zürich, Olga-Mayenfisch Stiftung, Hartmann-Müller Stiftung, Klaus Hermann Stiftung, and the Jubiläumsstiftung Swiss Life to MF; a fellowship of the Novartis Foundation, formerly Ciba-Geigy Jubilee Foundation to VVW. The SHCS drug resistance database is supported by the SHCS research foundation. The Zurich HIV Transcription study was supported by an unrestricted research grant from Abbott, Switzerland. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: HFG has served as a consultant and medical advisor for Abbott Laboratories, Bristol-Myers Squibb, Boehringer Ingelheim Pharmaceuticals, Gilead Sciences, GlaxoSmithKline, Pfizer, Tibotec Therapeutics, and ViiV Healthcare and has received unrestricted research and educational grants from Abbott Laboratories, Bristol-Myers Squibb, Gilead Sciences, Merck Sharp & Dohme, and Pfizer. No company was involved in the study design, in generation, analysis and interpretation of any data. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials.

* E-mail: marek.fischer@usz.ch (MF); huldrych.guenthard@usz.ch (HFG)

• These authors contributed equally to this work.

¶ These authors also contributed equally to this work.

Introduction

Current combination antiretroviral therapy (cART), despite its potency in suppressing active viral replication [1,2] and its power in reducing mortality and morbidity of HIV-1 infection [3,4], has not resulted in eradication or induction of treatment-free periods of remission of HIV-1 replication [5,6,7]. A pool of HIV-1 infected long-lived latently infected memory T-lymphocytes has been reported to be the major reservoir that confers HIV-1 infection resilient to eradication [8,9,10]. The frequency of latently infected cells was reported to range from 1 to 20 cells per 10^6 resting $CD4^+$ T-cells [9]. However, as these assays rely on technically demanding ex-vivo outgrowth assays, their results likely underestimate the size of the latent reservoir. Recent studies used PCR-based methods, in which latency had been defined as active viral transcription in the absence of viral progeny production. The resulting estimates of the size of the peripheral blood latent reservoir were at least 5–10 times higher in resting $CD4^+$ T-cells [11] as well as in total PBMC [12] when compared to viral outgrowth assays [9]. Apart from cells in peripheral blood, other

Own contributions

Benefits of early treatment in acutely HIV-1 infected patients remain controversial. Despite considerable side effects and also costs, which need to be taken into account in relation to decades of expected treatment time, early treatment may prevent expansion of the latent reservoir as well as viral escape to HIV-1 specific immune responses as viral diversity remains low. Thus, in this study, intra- and extracellular viral dynamics were assessed before, during, and after treatment of acute HIV-1 infection. For this purpose, HIV-1 RNA transcription patterns in PBMC and levels of plasma viremia were monitored. Comparison with a group of patients, who had initiated combination antiretroviral therapy (cART) during chronic infection, was performed cross-sectionally.

My contributions to this study were based on the quantitative PCR (qPCR) assays described in chapter 2 (Althaus *et al*., 2010). Measuring HIV-1 RNA transcription patterns in PBMC isolates of HIV-1 infected patients requires single copy sensitive qPCR, especially in patients under cART. Therefore, different primer and FH-probe combinations published in the previous paper were used in this study. Additionally, because of the high variability of the HIV-1 genome, I assisted in the search for optimal patient matched, i.e., sequence-adapted primers and FH-probes by applying the established rules regarding both target conservation and physical properties, such as melting temperatures, baseline fluorescence and secondary structure. Furthermore, I supervised the qPCR procedures, and helped in analyzing and interpreting the qPCR data.

Single copy sensitive assays were essential for this study as one major finding was that unspliced RNA was detectable during cART in 42% of patients with cART initiated during acute HIV-1 infection as opposed to 87% of patients, who initiated cART during chronic HIV-1 infection; a statement that could not have been made without highly sensitive assays. This work indicates that early intervention results in profound depletion of PBMC expressing HIV-1 RNA, which is contrary to chronically infected patients on cART.

Profound Depletion of HIV-1 Transcription in Patients Initiating Antiretroviral Therapy during Acute Infection

Adrian Schmid*, Sara Gianella*, Viktor von Wyl, Karin J. Metzner, Alexandra U. Scherrer, Barbara Niederöst, Claudia F. Althaus, Philip Rieder, Christina Grube, Beda Joos, Rainer Weber, Marek Fischer, Huldrych F. Günthard

University of Zürich, University Hospital Zürich, Division of Infectious Diseases and Hospital Epidemiology, Zürich, Switzerland
*These authors contributed equally to the manuscript.

Background: Although combination antiretroviral therapy (cART) initiated in the acute phase of HIV-1 infection may prevent expansion of the latent reservoir, its benefits remain controversial. In the current study, HIV-1 RNA transcription patterns in peripheral blood mononuclear cells (PBMC) were monitored during acute cART to assess the effect of early treatment on cellular viral reservoirs.

Methodology/Principal Findings: Acutely HIV-1 infected patients (n=24) were treated within 3-15 weeks after infection. Patients elected to cease treatment after ≥1 year of therapy. HIV-1 DNA (vDNA), HIV-1 RNA species expressed both in latently and productively infected cells, unspliced (UsRNA), multiply spliced (MsRNA-tatrev; MsRNA-nef), and PBMC-associated extracellular virion RNA (vRex), expressed specifically by productively infected cells, were quantified in PBMC by patient matched real-time PCR prior, during and post cART. In a matched control-group of patients on successful cART started during chronic infection (n=15), UsRNA in PBMC and vDNA were measured cross-sectionally.

In contrast to previous reports, PBMC-associated HIV-1 RNAs declined to predominantly undetectable levels on cART. After cART cessation, UsRNA, vRex, and MsRNA-tatrev rebounded to levels not significantly different to those at baseline (p>0.1). In contrast, MsRNA-nef remained significantly lower as compared to pretreatment (p=0.015). UsRNA expressed at the highest levels of all viral RNAs, was detectable on cART in 42% of patients with cART initiated

during acute infection as opposed to 87% of patients on cART initiated during chronic infection (Fisher's exact test; p=0.008). Accordingly, UsRNA levels were 105–fold lower in the acute as compared to the chronic group.

Conclusion: Early intervention resulted in profound depletion of PBMC expressing HIV-1 RNA. This is contrary to chronically infected patients who predominantly showed continuous UsRNA expression on cART. Thus, antiretroviral treatment initiated during the acute phase of infection prevented establishment or expansion of long-lived transcriptionally active viral cellular reservoirs in peripheral blood.

Introduction

Current combination antiretroviral therapy (cART), despite its potency in suppressing active viral replication [1,2] and its power in reducing mortality and morbidity of HIV-1 infection [3,4], has not resulted in eradication or induction of treatment-free periods of remission of HIV-1 replication [5-7]. A pool of HIV-1 infected long-lived latently infected memory T-lymphocytes has been reported to be the major reservoir that confers HIV-1 infection resilient to eradication [8-10]. The frequency of latently infected cells was reported to range from 1 to 20 cells per 10^6 resting $CD4^+$ T-cells [9]. However, as these assays rely on technically demanding ex-vivo outgrowth assays, their results likely underestimate the size of the latent reservoir. Recent studies used PCR-based methods, in which latency had been defined as active viral transcription in the absence of viral progeny production. The resulting estimates of the size of the peripheral blood latent reservoir were at least 5-10 times higher in resting $CD4^+$ T-cells [11] as well as in total PBMC [12] when compared to viral outgrowth assays [9]. Apart from cells in peripheral blood, other sites such as lymphoid tissues [13-15], the gastrointestinal tract [16-19], the brain [20], and the genital tract [21], have been reported to contribute to latent viral sanctuaries.

Due to the low expression levels of viral RNA in latently infected cells [12,14,22-24] and viral antigens, the latent viral reservoir is greatly inaccessible for adaptive and innate immune defenses. It has been proposed that local bursts of viral replication are initiated by immune activation in response to specific antigens [25-27] or due to random self-activation by the viral transactivator Tat [28]. In the presence of potent cART, such bursts will be dead-ended, because the viral particles produced may not initiate new rounds of infection. Conversely, after cessation of cART, this process will lead to rapid viral rebound [29] initiated stochastically from single or oligoclonal archival proviruses [30].

Nevertheless, the canon that HIV-1 replication, as measured by levels of plasma viremia, will inevitably proceed in the absence of cART [31,32] or recur soon after its cessation [29,33] has been challenged by important exceptions. Due to strong and specific immunological responses, so called elite-controllers contain viral replication to levels below a clinically relevant threshold (<50 copies/ml) in the absence of cART (reviewed in [34]). Furthermore, in some instances treatment initiated during the acute or early phases of HIV-1 infection had resulted in control of viremia after treatment

cessation [35-38]. These anecdotal cases have been used as precedents supporting early treatment. In addition, two rationales have been proposed to affirm possible benefits of early treatment of HIV-1, despite considerable side effects [39-41] as well as costs [42,43], which need to be considered in relation to decades of expected treatment time.

Paul Ehrlich's paradigm to "hit early and hard" in treatment of infectious disease [44] to limit spread of an infectious pathogen and to contain its population size, toxicity, and its potential to escape immunological and chemotherapeutic/medical control [45] is still valid and accepted. This concept was substantiated for HIV-1 by Strain et al., who showed that the size of latent reservoirs was smaller in patients with treatment initiation in the acute phase than in those who initiated cART during chronic infection [46]. In addition to that, the notion that acute HIV-1 infection is generally associated with oligoclonal viral quasispecies of low phylogenetic diversity [47,48] led to the hypothesis that by early treatment viral escape to specific HIV-1 antibody- and cytotoxic T-cell responses may be avoided or deferred resulting in clinical benefit and/or virological control of the virus. This concept was corroborated by a study showing that transmission of multiple quasispecies was associated with faster disease progression than mono- or oligoclonal primary infection [49]. Low viral diversity has also been shown to be associated with modest but significant control of viremia in chronic infection [50]. Moreover, superinfection can lead to the loss of control of viremia [51-56].

To elucidate these issues in the current study, intra- and extracellular viral dynamics were assessed before, during, and after treatment of acute HIV-1 infection. For this purpose HIV-1 RNA transcription patterns in PBMC and levels of plasma viremia were monitored. Comparison with a group of patients, who had initiated cART during chronic infection, was performed cross-sectionally. Unexpectedly, cell-associated viral RNA burden in PBMC became mostly undetectable in the acutely infected patient group. The levels of UsRNA, the most abundant cellular HIV-1 RNA *in vivo*, were at least one order of magnitude lower during cART initiated in acute than in chronic infection. These results strongly indicate that patients initiating cART in the acute phase of infection may attain a substantial virological benefit from early intervention.

Results

Kinetics of viral RNA levels in response to treatment and cessation of cART

Acutely HIV-1 infected patients (n=24) were treated with cART within 3-15 weeks after infection. Plasma viremia was monitored using standard diagnostic quantitative PCR-assays (qPCR) and PBMC-associated HIV-1 nucleic acids (figure 1) were quantified by patient matched qPCR [11,12,57,58]. HIV-1 DNA (vDNA) was assessed as a measure of total infected cells, comprising latently and productively infected cells as well as cells harboring defective genomes.

To investigate the activity of these HIV-1 infected cells, distinct PBMC-associated viral RNA species (vRNAs) were quantified: unspliced RNA (UsRNA) and 2 types of multiply spliced encoding tat and rev (MsRNA-tatrev) or nef (MsRNA-nef). These RNAs were found to be expressed both in latently [12,23,59] and productively infected cells, however, at much higher levels in the latter cell population. As a marker for productively infected cells, PBMC-associated extracellular virion RNA (vRex) was measured [11,12,14,22,57].

Initiation of cART resulted in a decay of viral RNA production (figure 2 A/B left panel). Whereas plasma viremia followed kinetics which were consistent with a 2-3 phasic exponential decay, vRex dropped to undetectable levels almost instantly after initiation of cART. Decay of UsRNA and MsRNAs was intermediate between that of plasma viremia and vRex (data not shown). Upon cessation of cART, rebound of HIV-1 nucleic acids occurred quickly in plasma (figure 2A, right panel), but was apparently delayed in PBMC (figure 2B, right panel, figure 3B).

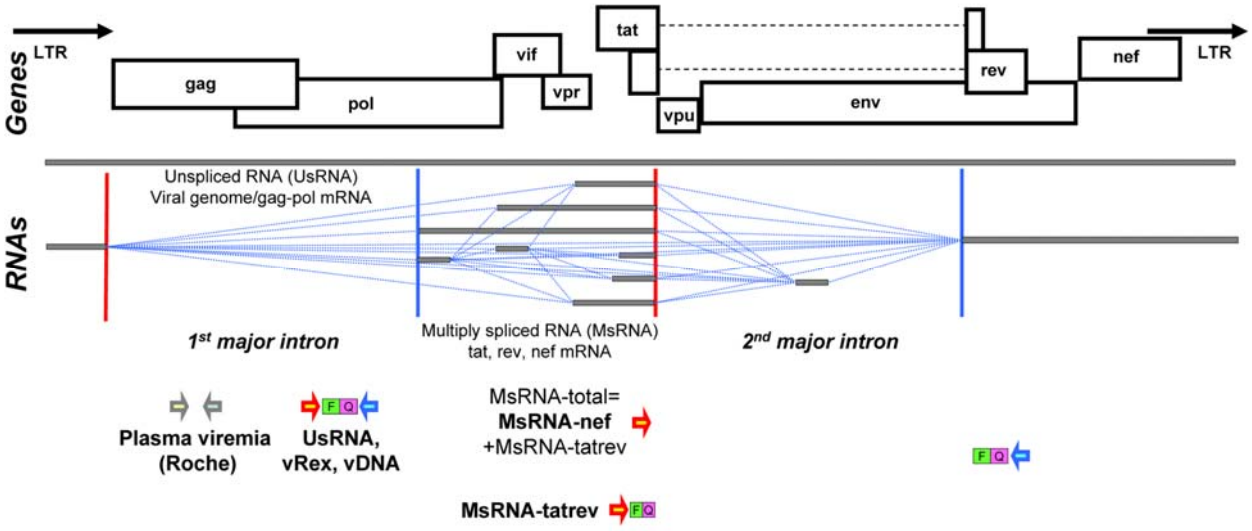

Figure 1: Map of HIV-1 amplica for qPCR assays. Splice acceptors and donor sites are shown by red and blue vertical lines, respectively. Exons are depicted in grey bars. Blue dotted lines show documented or predicted splice events. Sense primers used in qPCR assays are depicted by red arrows antisense/cDNA primers by blue arrows. Fluorescent hydrolysis probes [58] (FH-probes / TaqMan®-probes) are shown by green/pink bars. Assays for MsRNAs share a common antisense/cDNA primer and in some instances, when a probe for MsRNA-tatrev was not available, a common FH-probe 3' of the splice-acceptor of the 2nd major intron. Note that this map does not show mRNAs encoding Env because these transcripts were not assessed in the current study.

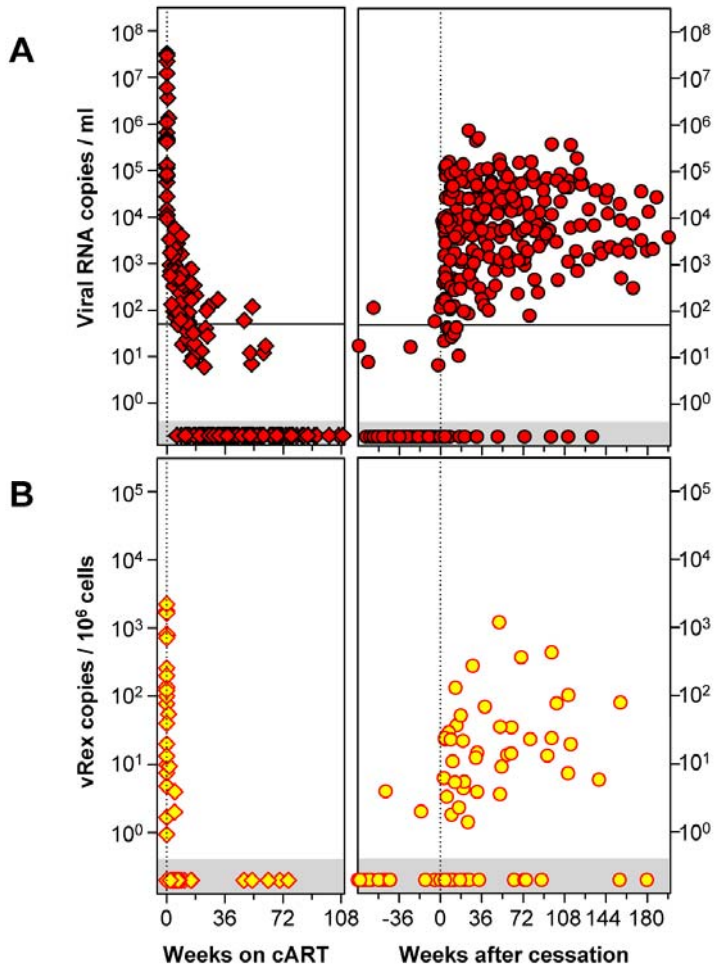

Figure 2: Longitudinal course of HIV-1 virion production. Levels of plasma viremia (**A**) and of PBMC-associated virions, vRex (**B**) of 24 acutely infected patients, were plotted against time after initiation of cART (0 weeks, left panels) or against time after cART cessation (defined as 0 weeks in the right panels). Note that some data points are depicted in the left panels are also shown in the right panels to facilitate visualization of the effects of cART cessation. The amount of vRex measurements is lower, because cell sampling was done less frequently than plasma HIV-RNA was measured. Data within the grey horizontal areas show PCR-negative samples below the detection limit of the assays applied. The black line in panel A depicts the clinically used threshold for plasma viremia of 50 RNA copies/ml.

Time to event analysis of HIV-1 RNA detection during and after cART

To corroborate these observations and to extend them to UsRNA and MsRNAs, empirical distributions of the proportions of patients with detectable viral RNAs after initiation of cART were plotted (figure 3A). Similarly, distribution of the proportion of patients experiencing viral rebound after therapy cessation was addressed (figure 3B). Upon initiation of cART, disappearance of vRex took a median time of 4.5 weeks (quartiles 4, 6 weeks). Taking into account that the second sample following baseline in the current study was obtained at a median time of 4.4 weeks (quartiles 4, 5 weeks) depletion of vRex occurred instantaneously after cART initiation. MsRNAs declined significantly later to undetectable levels (Mann-Whitney, p=0.002), namely within a median time of 8.4 weeks (quartiles 5, 13 weeks). The next parameter to reach its detection limit with a significant delay compared to MsRNA (Mann-Whitney, p=0.003) was UsRNA with a median time to the first undetectable measurement of 13.5 weeks (quartiles 9, 47 weeks). Finally, plasma viremia dropped to undetectable levels within a median time of 24.6 weeks (quartiles 12, 44 weeks), which was not significantly different to the decay of UsRNA (Mann-Whitney, p=0.52). In general, after detection of the first undetectable HIV RNA value, all RNA species measured remained undetectable at almost all timepoints measured during treatment.

After cessation of cART, plasma viremia reached levels above the threshold of 50 copies/ml in a median time of 4.2 weeks (quartiles 3, 10 weeks). Rebound of UsRNA and MsRNA proceeded with some delay without reaching statistical difference (Mann-Whitney, p=0.30 and p=0.07, respectively): Median time to rebound was 8.0 weeks (quartiles 4, 12 weeks) and 9.8 weeks (quartiles 4, 19 weeks) for UsRNA and MsRNA, respectively. Lastly, within a median time of 17.9 weeks (quartiles 8, 56 weeks) vRex rose to detectable levels. Its rebound was statistically different from that of plasma viremia and of UsRNA (Mann-Whitney, p=0.007 and 0.02, respectively) but not from MsRNA (Mann-Whitney, p=0.06). It has to be noted that vRex was measured less frequently than plasma or PBMC viremia. Therefore the median time to rebound may have been overestimated.

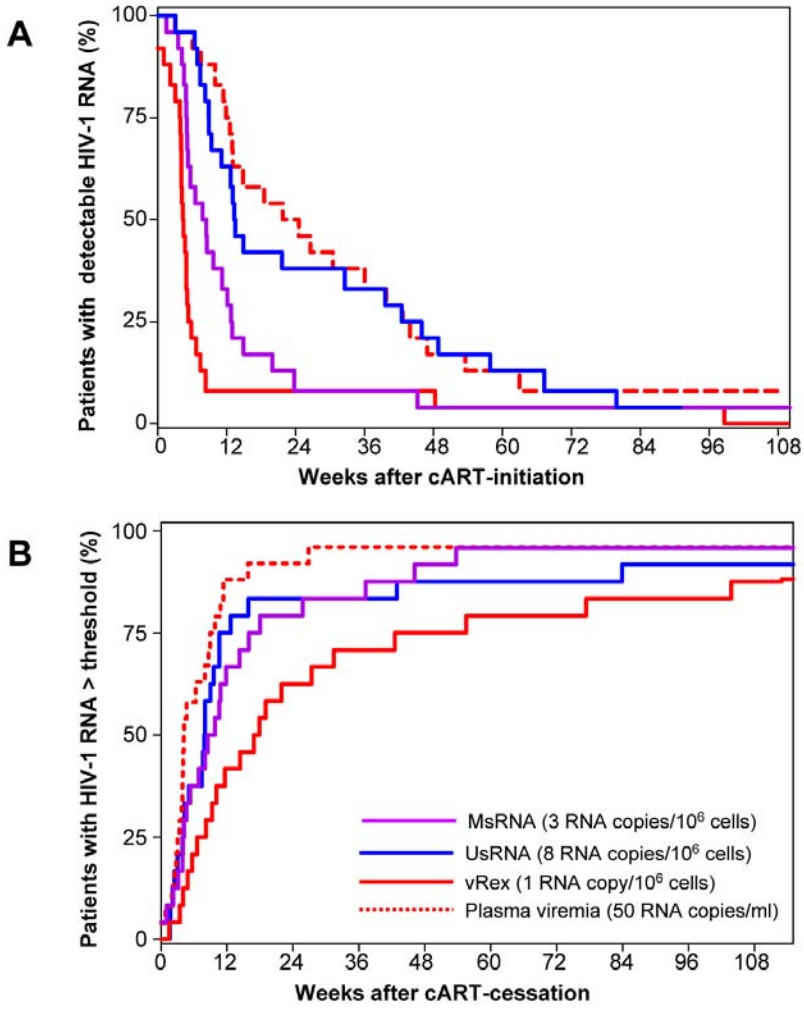

Figure 3: Empirical distribution of times to virological end-points. Time to event analysis showing the percentage of patients before reaching their first PCR-negative measurement after initiation of cART (**A**) or reaching measurable viral RNA levels above a defined threshold after cessation of cART (**B**). Thresholds are indicated in parentheses in panel B and were defined as approximately 3-fold elevation over mean on cART levels for cellular HIV-1 RNAs. Threshold of plasma viremia was set at 50 copies/ml. On average vRex (red lines), MsRNAs (magenta), UsRNA (blue) and plasma viremia (red broken lines) dropped to undetectable levels 4.5, 8.4, 13.5, and 24.6 weeks after initiation of cART, respectively (A) and rebounded within 17.9, 9.8, 8.0, and 4.2 weeks after treatment cessation (B), (dotted lines). Note that for analysis of plasma viremia only time-points were considered for which also UsRNA and MsRNA measurements had been available.

Magnitude of viral nucleic acid levels at baseline and post-cessation

To investigate whether cART initiated in acute infection had an impact beyond its cessation, the levels of HIV-1 nucleic acids at baseline and after treatment cessation were compared. Maxima were used to represent post-cessation HIV-1 nucleic acid levels, to account for the variability among the different parameters regarding the time to reach stable set-points.

Plasma viremia (figure 4A) was significantly reduced after treatment cessation as compared to baseline (Wilcoxon signed rank test, p=0.002). This reduction can be attributed in part to the early treatment effect of cART (Gianella, von Wyl *et al*., unpublished data), but also to a naturally occurring "first peak" of viremia observed in acute HIV-1 infection [60]. We observed a less pronounced impact of early cART on cellular viral nucleic acid levels after treatment cessation. Only MsRNA-nef showed persistent significant reduction after cART (Wilcoxon signed rank test, p=0.02), whereas the remaining parameters, vDNA, vRex, UsRNA, and MsRNA-tatrev (figure 4B-E) tended to be lower after treatment stop compared to baseline but without reaching statistical significance (Wilcoxon signed rank test, p>0.10).

Viral reservoirs during cART

As opposed to baseline and post-cessation viral nucleic acid levels, which were represented by single measurements, on-cART levels were estimated by calculating means. Taking into account the rapid decay of vRex, the mean of samples at time-points ≥2 weeks after initiation of treatment was used. For all other parameters measurements taken at time points ≥24 weeks were used for calculation of mean values. The rationale for choosing this time-window was that the parameter with the slowest decay kinetics, plasma viremia, reached undetectable levels within a median time of around 24 weeks as shown in figure 3A. Accordingly, the resulting mean values for plasma viremia on cART were well below the clinically used threshold of 50 copies/ml (figure 4A), except for one patient (#133; 59 copies/ml).

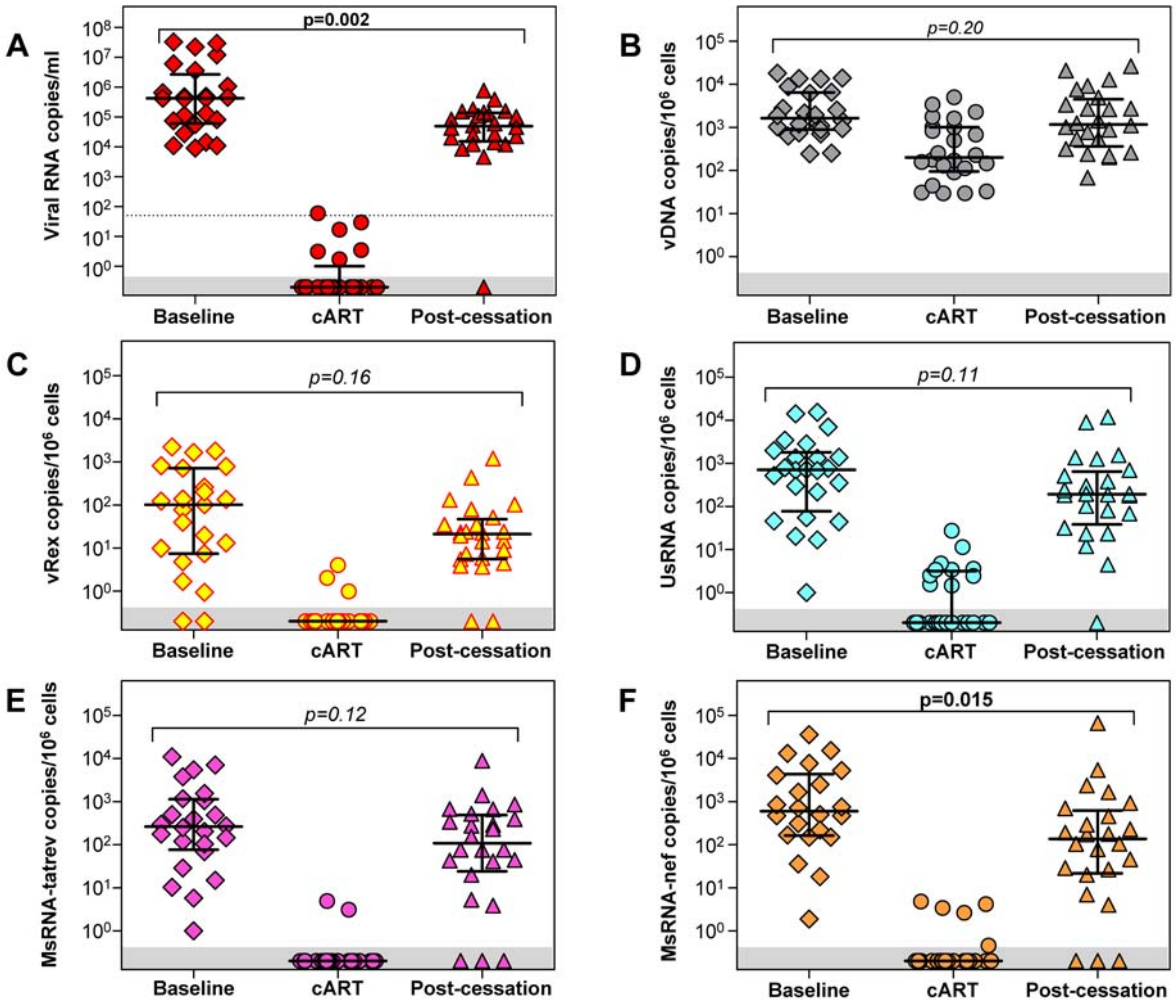

Figure 4: Virological parameters before, during, and after cART. Viral nucleic acid levels at baseline (diamonds), and average values during the time windows defined as on cART (circles, vRex ≥2 weeks, plasma viremia, PBMC-associated UsRNA and MsRNAs ≥24 weeks) and post-cessation maximal values (triangles). Data-points within the grey horizontal areas show PCR-negative samples below the detection limit of the assays applied. P-values within panels show significance levels (Wilcoxon signed rank test) between baseline and post-cessation. Bars show medians and quartiles. The dotted line in panel A depicts the clinically used threshold for plasma viremia of 50 RNA copies/ml.

HIV-1 DNA persisted at a level of 823±253 copies/10^6 PBMC (mean±sem) during cART initiated in acute infection. However, proviruses in these infected cells were transcriptionally almost completely silent. Expression of MsRNAs and vRex was mostly undetectable and attenuated to mean levels <1 copy/10^6 PBMC (MsRNA-tatrev: 0.34±0.24, MsRNA-nef: 0.65±0.30, vRex: 0.29±0.19; mean±sem). Only expression of UsRNA was occasionally detected at a mean level of 2.6±1.2 copies/10^6 PBMC. The absence of vRex during cART observed in the current study was similar to findings of previous studies [11,12,14,22] indicating a fast and complete loss of productively infected cells as a consequence of cART. However, the almost complete depletion of MsRNA and the profound reduction of UsRNA were highly unexpected, as we and others have shown their presence at measurable levels in latently infected cells of HIV-1 infected patients on cART [11,12,59].

Comparison of the effects of cART initiated in primary versus chronic infection

To investigate whether the observed depletion of viral transcription could be attributed to the early initiation of cART, PBMC obtained from patients, who had initiated antiretroviral therapy in the chronic phase of HIV-1 infection, were assessed for the presence of viral RNA in PBMC. Latently infected cells expressing solely UsRNA had been shown to persist during cART at the highest frequency among all cell classes expressing viral RNA and UsRNA had been consistently detected at the highest levels during cART in previous studies [11,12,57,61]. Therefore, UsRNA and vDNA were quantified in patients in whom cART had been initiated during chronic infection (n=15, for patient characteristics see Table S1) using identical conditions as for the acute patient group, including RNA extraction, use of patient matched primers, and PCR input volumes. Levels of plasma viremia in the chronic control group were <50 copies/ml (mean: 10 copies/ml, range 0-28) and treatment had lasted for a median time of 50 weeks (36-183 weeks). In this group of patients, 86% of specimens were positive for UsRNA, which is significantly higher than the percentage of patients in the acute group showing any UsRNA positive PCR on cART (42%, Fisher's exact test, p=0.008). Furthermore, the average level of UsRNA in the chronic control group (mean=272 copies/10^6 cells, range 0-1486) was 105-fold higher than in the acute group (Mann-Whitney, p<0.0001, figure 5A).

Levels of vDNA were significantly elevated (Mann-Whitney, p=0.04) in the chronic group as compared to the patients treated during acute infection. To adjust for this

difference in the total number of infected cells, the average viral transcriptional rate i.e., the ratio of viral RNA to vDNA, was calculated [11]. Similarly as the absolute levels of UsRNA, viral transcriptional rates were elevated with high statistical significance (12-fold, Mann-Whitney, p=0.0002) in the chronic group as compared to patients in the acute group (figure 5B).

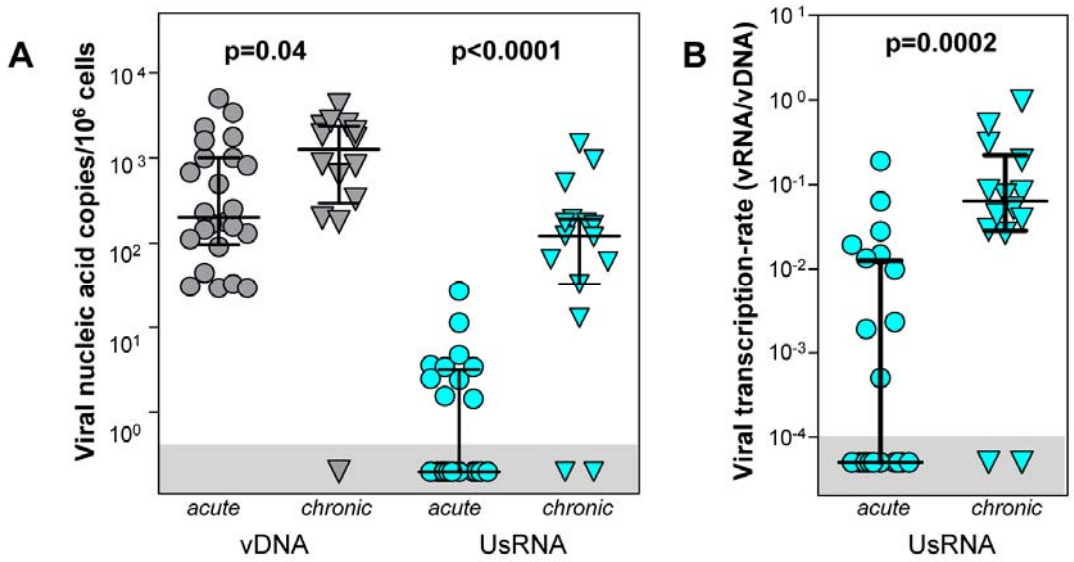

Figure 5: Viral nucleic acid expression in acute and chronic patients during cART. Mean levels of vDNA (grey symbols) and UsRNA (blue symbols) (**A**) or viral transcription rates (**B**) of patients that had initiated cART during acute infection (acute, n=24, circles) were assessed during the on-cART time window (≥24 weeks) and compared to single measurements in patients that had initiated cART during chronic infection (chronic, n=15, triangles, median treatment time 50 weeks, range 36-183, see Table S1 for details). Data-points within the grey horizontal areas show PCR-negative samples below the detection limit of the assays applied. P-values within panels show significance levels (Mann-Whitney test) between the acute and the chronic group. Bars show medians and quartiles.

Discussion

The main aims and rationales for antiretroviral therapy in acute HIV-1 infection are confinement of genetic diversity of viral quasispecies to prevent escape from adaptive immunity, restriction of $CD4^+$ T-cell loss, preservation of immune functions, and inhibition of initial viral spread throughout the possible sites of replication. Despite considerable efforts from various research groups, to date evidence from clinical studies in humans to empirically substantiate these concepts remains limited. Here we performed a longitudinal in-depth analysis using patient matched highly sensitive qPCR assays to detect different HIV-1 gene trancripts to assess effects of early treatment on cellular reservoirs of HIV-1 in peripheral blood.

HIV-1 transcription was monitored longitudinally before, during, and after cessation of treatment of acute HIV-1 infection in well characterized patients [60,62]. This analytical approach was applied because numerous studies revealed persistence of cellular HIV-1 RNA for years in patients on cART despite suppression of plasma viremia [14,15,17,57,61,63-65]. By choosing a patient matched approach for qPCR with single copy sensitivity, it was anticipated that not only levels of UsRNA, but also the less abundant multiply spliced RNAs could be quantified and used to assess the viral transcription pattern of the PBMC reservoir of HIV-1 during cART [11,12,57].

The major finding of this study was that early cART initiated during acute HIV-1 infection significantly depleted the number of transciptionally active proviruses by an order of magnitude when compared to levels detected in patients treated during chronic infection.

Specifically, productively infected PBMC, hallmarked by expression of cell attached virions (vRex) were depleted within a median time of 4.5 weeks after cART initiation. The remaining $HIV\text{-}1^+$ PBMC, expressing solely intracellular viral RNA, can therefore be viewed as latently infected [11,12,14]. These cells faced depletion within a median time of 24 weeks. This resulted in >100-fold reduction of viral transcription levels and at least 10-fold lower average viral transcriptional rates as compared to patients who had started cART during chronic infection. It is unlikely that the strong reduction of viral transcription in PBMC in primary HIV infected patients was due to the potency of treatment rather than to timing of antiretroviral therapy because the control group of chronically infected patients received similar treatment.

Decay of HIV-1$^+$ cells in response to treatment occurred in a significantly staggered mode according to their pattern of viral RNA expression and life-span as previously described [12]. Productively infected cells expressing vRex vanished almost instantly after initiation of treatment, then cells expressing MsRNA followed with a delay of about 4 weeks, ultimately a further 5 weeks later, cells expressing solely UsRNA pursued. Thus, even HIV-1 RNA expressing cells with the longest life spans [12] approached depletion.

These findings imply that early cART led to the clearance of long-lived cells harboring transcriptionally active latent proviruses. In an early study, Markowitz and colleagues showed persistence of viral transcription in PBMC when antiretroviral therapy was initiated within 90 days after onset of primary symptoms [17]. This discrepancy can most likely be explained by a 5-6 week earlier initiation of therapy in the present study. Of note, our findings are in agreement with the observations of Strain and colleagues [46]. In their study, using viral outgrowth assays of resting CD4$^+$ T-lymphocytes, a cohort of patients on early cART showed lower levels of latently infected cells than a control group with deferred cART initiation. In addition, two small studies in humans and in the SIV model also suggest that HIV-1 suppression as measured by unspliced HIV-1, respectively SIV RNA can be reduced by treatment during primary HIV infection in the gut associated lymphoid tissue [66,67].

After cessation of cART, virological parameters reappeared inverted to their decay. UsRNA rose to significant levels after 8 weeks, followed two weeks later by MsRNA and finally vRex reappeared with a significant delay. The observations that viral rebound in plasma preceded that of cellular viral RNAs and plasma viremia was the last parameter to decay during cART, implies that PBMC can be viewed as independent compartment to some degree separate from the source of virus appearing in plasma. These dynamics potentially suggest that repopulation of the PBMC compartment of rebounding transcriptionally active PBMC was most likely due to "*de novo*" infection of PBMC by plasma virions rather than by reactivation of latently infected PBMC. The rebounding virus most likely originates from distinct compartments such as the secondary lymphoid organs (e.g., gut associated lymphoid tissue (GALT), lymph nodes, spleen), the cerebral nervous system, the kidneys, or the genital tract. Very recent data by Yukl *et al.* [19] suggests that most likely more than 83% of all HIV-1 infected cells under therapy are to be found in the GALT and that the RNA/DNA ratio on treatment tends to be higher in this

compartment than in the peripheral blood. Furthermore, GALT is continuously exposed to bacterial antigenic stimuli from gut commensals. In aggregate, the high HIV burden together with continuous antigenic stimulation makes this large reservoir a prominent candidate as a source for the rebounding virus. Of note, the RNA/DNA ratios of GALT and PBMC in the study by Yukl was measured with exactly the same methods as in our study.

The fact that post-cessation control of viral RNA expression was achieved in several patients during the post treatment observation period supports this hypothesis: Plasma viremia was undetectable in patient #72, control of UsRNA expression was experienced in patient #25, control of MsRNA expression was achieved by patients #25, #56, and #92, and control of vRex was experienced by patients #72 and #99.

The observation that MsRNA-nef levels, a parameter reported to be associated with viral latency [57,68] were significantly reduced as compared to baseline further indicates that the latent reservoir was reduced by initiation of cART during acute infection. Conversely, it cannot be fully excluded that the decreased levels of MsRNA-nef after therapy cessation may be explained by the “first peak” of viral replication in acute HIV-1 infection and its subsequent spontaneous decay independent of cART [69].

Taken together, our data demonstrate a profound virologic effect of treatment initiation during acute HIV-1 infection. The unprecedented finding that residual viral transcription was almost completely depleted, suggests that the latent reservoir of HIV-1 during cART can be reduced by early intervention and that proof of concept studies aiming at HIV-1 eradication or remission of HIV-1 replication should be initiated in patients during acute infection.

Materials and Methods

Ethics Statement:

In accordance with the guidelines of the Ethics Committee of the University Hospital of Zurich, written informed consent was obtained from all participants of the studies analyzed: the Zurich Primary HIV Study (ZPHI), the Swiss Spanish Intermittent Therapy Trial (SSITT), and the Zurich HIV-1 Transcription Study.

Patients and specimens

Patients with treatment initiated in the acute phase of HIV-1 infection (acute group). Patients (n=24; Table S1) included in our study were enrolled in the Zurich Primary HIV Study (ZPHI, http://clinicaltrials.gov, NCT00537966) [62] a substudy of the Swiss HIV Cohort Study (SHCS). All 24 subjects had a documented acute or recent primary HIV-1 infection at presentation. Estimation of time after infection was as described by Rieder et al [60]. Patients were included in the analysis only if sequences were available to match primers in *pol, tat*, and *env*.

Blood samples were collected at the day of the very first cART application (=baseline sample), longitudinally during treatment, and after treatment cessation. EDTA-blood samples were collected and separated into cells and plasma according to standard procedures [22]. Cellular material was split into aliquots, each comprising 2 million cells, and finally stored at -80°C as dry cell pellets. Sampling of specimens for analysis of plasma viremia and cellular nucleic acids was performed as described in Methods S1. All patients were extensively tested for genotypic drug resistance: Only one patient showed the D67N mutation at baseline in the routine genotyping test and in addition, one patient showed 0.02% K103N and one 2.4% M184V harboring minority species [62].

Patients with treatment initiated in the chronic phase of HIV-1 infection (chronic group). PBMC samples for a cross-sectional control group of 15 patients treated in the chronic phase after seroconversion were obtained from the Swiss Spanish Intermittent Therapy Trial (SSITT) [33] at the last time-point before treatment cessation. In addition, specimens collected for The Zurich HIV-1 Transcription Study (INFZ VTA 02.00 [12]) were analyzed (n=9). As in the acute group, patients in the chronic group participated also in the Swiss HIV Cohort Study (SHCS) and HIV-1 *pol*-

sequences were available from the SHCS genotypic drug resistance database [70]. Patient characteristics of the chronic group are listed in Table S1.

Nucleic acid extraction/preparation

Total PBMC RNA was extracted with the RNeasy extraction kit on the 'QIAcube' extraction device (Qiagen Hombrechtikon, Switzerland) using an initial volume of 0.35 ml for cell lysis, DNAse I digestion on the extraction column and a final elution volume of 0.1 ml.

Selective isolation of PBMC-associated extracellular, virion-encapsidated genomic HIV-RNA, vRex, free of intracellular nucleic acid, was performed as described [14,22,71] using an elution volume of 0.1 ml. Preparation of protease digested total cell-lysates for DNA quantification was performed according to Christopherson et al. [72] with minor modifications [11].

Design of HIV-1 patient matched qPCR

To ensure accurate quantification a patient matched approach reaching single copy sensitivity was used for qPCR measurements of RNA and vDNA [11,12,57,58] using specifically designed fluorescent hydrolysis probes [58]. The location of qPCR primers and probes is shown in figure 1; sequences are outlined in detail in Table S2 and Methods S1. Individual sequence information to adjust qPCR assays for UsRNA, vRex, and vDNA, was obtained from the SHCS genotypic drug resistance database (HIV-1 *pol*-sequences) [70].

To obtain sequence information to match qPCR primers and probes for MsRNAs of predominant quasispecies, two regions were amplified from each patients total DNA by a nested PCR scanning approach as described [11] (positions 5833-6152 and 7992-8567 in the HXB2 genome, GenBank accession number K03455).

As previously described [11,12,57] two different assays for MsRNA were performed: MsRNA-total and MsRNA-tatrev from which the MsRNA-nef was calculated using the difference in copy numbers between the two measurements. To avoid redundancy, MsRNA-total was not reported. Thus, only MsRNA-tatrev and MsRNA-nef are shown in the current data-set and used for statistical calculations. In 11% of measurements, it occurred that copy numbers of MsRNA-total were nominally lower or equal to MsRNA-tatrev. Calculation of MsRNA-nef was considered as not applicable in these cases.

Quantitative real-time PCR measurements

Assays for RT-qPCR were performed as described [57,73] using an ABI 7500 real-time thermocycler (Applied Biosystems, Rotkreuz, Switzerland) with 0.005µM ROX as passive reference, 0.2 µM fluorescent probe and 5 µl RNA as template in a final volume of 35 µl aqueous phase plus 15 µl paraffin. Quantification of vDNA was performed as described by Althaus *et al.* [58] in an IQ5 real-time thermocycler (Biorad, Basel, Switzerland) in a volume of 60 µl with 10 µl DNA template using HotStarTaq master mix (QIAGEN, Hilden, Germany) supplemented with PCR primers (1 µM each), probe (0.3 µM), and additional $MgCl_2$ (1.5 µM) by incubation for 15 min at 95°C and 60 cycles of 10 s at 95°C, 5 s at 55°C and 40 s at 60°C.

Calculation of HIV-1 nucleic acid copy numbers from qPCR data

Patient-nucleic acids were directly used to generate HIV-1 standard-curves. To obtain standard curves with maximal dynamic range, baseline specimens were used as standards for vDNA-measurements. Post-cessation time points with high plasma viremia and high vDNA were chosen. Standards were prepared as serial 3-fold dilutions in quadruplicates. Specimens other than the ones chosen as standards were measured in duplicate.

Linear regression used in standard real-time PCR analysis was applied for calculation of HIV-1 copy numbers:

Equation 1) $Ct = I + S \times \log_{10}$ (*copy numbers*),

where *I* is the Y-axis intercept of the regression line, depicting the point where Ct equals one copy ($=10^0$) and *S* is its slope. *I* and *S* were determined as described in Methods S1. In cases where only one duplicate HIV-1 PCR was positive and/or calculation resulted in less than 1 nominal copy number per reaction, HIV-1 nucleic acid copy numbers were censored to 1 copy per PCR [11,12].

Normalization to cellular input

To control for input and/or differences in qPCR efficiencies in RNA PCR the expression of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) against a standard dilution series of in vitro transcribed GAPDH RNA was measured in duplicates as described previously [73] and in Methods S1. Linear regression, as applied by the ABI7500 software, was used to evaluate GAPDH measurements. Mean GAPDH-RNA copies per 10^6 cell equivalents were empirically determined in PBMC specimens with known cell numbers (2 x10^6 cells sample, n=310) resulting in a conversion factor of 2.85 x10^7 copies GAPDH-RNA/10^6 PBMC.

Cellular input for DNA was quantified by beta-actin PCR using primers and probe described in Methods S1. As quantification standard for cellular input, 15 million PBMC were lysed in 1 ml cell lysis buffer, serial two-fold dilutions were measured and cellular input per PCR was calculated by linear regression using the software provided with the IQ5-system. Both for vDNA and viral RNA, the mean of duplicate PCR measurements was normalized to cellular input and finally expressed as HIV-1 nucleic acid copies per 10^6 PBMC.

Statistical analysis

GraphPad Prism 5.0 software (GraphPad Software, San Diego, CA) was used for statistical analyses. $P<0.05$ was considered as the level of significance. Mann-Whitney testing was applied unless otherwise indicated for group comparisons. Fisher's exact test was used in analysis of contingency tables. No adjustment for multiple testing was applied.

Acknowledgments

We are grateful to all the patients participating in the Zurich primary HIV infection Study, in the Zurich HIV transcription study, and the Swiss Spanish Intermittent Therapy Trial. Further we want to acknowledge Barbara Hasse, Urs Karrer, Rolf Oberholzer, Elisabeth Presterl, Reto Laffer, Ulrich von Both, Klara Thierfelder, Dominique Braun, Yvonne Flammer, Markus Flepp, and Thomas Frey for their dedicated patient care; Friederike Burgener, Dominique Klimpel, and Herbert Kuster for excellent laboratory assistance; Christine Vögtli, Danièle Perraudin, and Ingrid Nievergelt for administrative support. Furthermore, we thank Jürg Böni for running the Zurich part of the Swiss HIV drug resistance database. We thank also all the staff of the SHCS clinical centers and the data center.

References

1. Gulick, R. M. et al., Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. N Engl J Med 337 (11), 734 (1997).
2. Opravil, M. et al., Effects of early antiretroviral treatment on HIV-1 RNA in blood and lymphoid tissue: A randomized trial of double versus triple therapy. J Acquir Immune Defic Syndr Hum Retrovirol 23 (1), 17 (2000).
3. Egger, M. et al., Impact of new antiretroviral combination therapies in HIV infected patients in Switzerland: prospective multicentre study. Swiss HIV Cohort Study. BMJ. 315 (7117), 1194 (1997).
4. Palella, F. J., Jr. et al., Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. N Engl J Med 338 (13), 853 (1998).
5. Stebbing, J., Gazzard, B., and Douek, D. C., Where does HIV live? N Engl J Med 350 (18), 1872 (2004).
6. Richman, D. D. et al., The challenge of finding a cure for HIV infection. Science 323 (5919), 1304 (2009).
7. Dahl, V., Josefsson, L., and Palmer, S., HIV reservoirs, latency, and reactivation: prospects for eradication. Antiviral Res 85 (1), 286 (2009).
8. Wong, J. K. et al., Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. Science 278 (5341), 1291 (1997).
9. Finzi, D. et al., Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science 278 (5341), 1295 (1997).
10. Chun, T. W. et al., Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proc Natl Acad Sci U.S.A. 94 (24), 13193 (1997).
11. Kaiser, P. et al., Productive Human Immunodeficiency Virus 1 Infection in peripheral Blood Predominantely Takes Place in CD4/CD8 double negative T Lymphocytes. J Virol 81, 9693 (2007).
12. Fischer, M. et al., Biphasic decay kinetics suggest progressive slowing in turnover of latently HIV-1 infected cells during antiretroviral therapy. Retrovirology 5, 107 (2008).
13. Zhang, L. et al., Genetic characterization of rebounding HIV-1 after cessation of highly active antiretroviral therapy. J Clin Invest 106 (7), 839 (2000).
14. Fischer, M. et al., Attenuated and nonproductive viral transcription in lymphatic tissue of human immunodeficiency virus type 1 infected patients on potent antiretroviral therapy. J Infect Dis 189 (2), 273 (2004).
15. Günthard, H. F. et al., Residual human immunodeficiency virus (HIV) Type 1 RNA and DNA in lymph nodes and HIV RNA in genital secretions and in cerebrospinal fluid after suppression of viremia for 2 years. J Infect Dis 183 (9), 1318 (2001).
16. Anton, P. A. et al., Multiple measures of HIV burden in blood and tissue are correlated with each other but not with clinical parameters in aviremic subjects. AIDS 17 (1), 53 (2003).
17. Markowitz, M. et al., The effect of commencing combination antiretroviral therapy soon after human immunodeficiency virus type 1 infection on viral replication and antiviral immune responses. J Infect Dis 179 (3), 527 (1999).

18. Mehandru, S. et al., Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract. J Exp Med 200 (6), 761 (2004).
19. Yukl, S. et al., Differences in HIV Burden and Immune Activation within the Gut of HIV+ Patients on Suppressive Antiretroviral Therapy. J Infect Dis, In press. (2010).
20. Barber, S. A. et al., Mechanism for the establishment of transcriptional HIV latency in the brain in a simian immunodeficiency virus-macaque model. J Infect Dis 193 (7), 963 (2006).
21. Craigo, J. K. et al., Persistent HIV type 1 infection in semen and blood compartments in patients after long-term potent antiretroviral therapy. AIDS Res Hum Retroviruses 20 (11), 1196 (2004).
22. Fischer, M. et al., Residual cell-associated unspliced HIV-1 RNA in peripheral blood of patients on potent antiretroviral therapy represents intracellular transcripts. Antiviral Therapy 7 (2), 91 (2002).
23. Lassen, K. G., Bailey, J. R., and Siliciano, R. F., Analysis of human immunodeficiency virus type 1 transcriptional elongation in resting CD4+ T cells in vivo. J Virol 78 (17), 9105 (2004).
24. Hermankova, M. et al., Analysis of human immunodeficiency virus type 1 gene expression in latently infected resting CD4+ T lymphocytes in vivo. J Virol 77 (13), 7383 (2003).
25. Grossman, Z. et al., Ongoing HIV dissemination during HAART. Nat Med. 5 (10), 1099 (1999).
26. Günthard, H. F. et al., Effect of influenza vaccination on viral replication and immune response in persons infected with human immunodeficiency virus receiving potent antiretroviral therapy. J Infect Dis 181 (2), 522 (2000).
27. Douek, D. C. et al., HIV preferentially infects HIV-specific CD4+ T cells. Nature 417 (6884), 95 (2002).
28. Weinberger, L. S., Burnett, J. C., Toettcher, J. E., Arkin, A. P., and Schaffer, D. V., Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell 122 (2), 169 (2005).
29. Fischer, M. et al., HIV RNA in plasma rebounds within days during structured treatment interruptions. AIDS 17 (2), 195 (2003).
30. Joos, B. et al., HIV rebounds from latently infected cells, rather than from continuing low-level replication. Proc Natl Acad Sci U S A 105 (43), 16725 (2008).
31. Galetto-Lacour, A. et al., Prognostic value of viremia in patients with long-standing human immunodeficiency virus infection. Swiss HIV Cohort Study Group. J Infect Dis 173 (6), 1388 (1996).
32. Michael, N. L. et al., Development of calibrated viral load standards for group M subtypes of human immunodeficiency virus type 1 and performance of an improved AMPLICOR HIV-1 MONITOR test with isolates of diverse subtypes. J Clin Microbiol 37 (8), 2557 (1999).
33. Fagard, C. et al., A prospective trial of structured treatment interruptions in human immunodeficiency virus infection. Arch Intern Med 163 (10), 1220 (2003).
34. Deeks, S. G. and Walker, B. D., Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. Immunity 27 (3), 406 (2007).
35. Lisziewicz, J. et al., Control of HIV despite the discontinuation of antiretroviral therapy. N Engl J Med 340 (21), 1683 (1999).

36. Rosenberg, E. S. et al., Immune control of HIV-1 after early treatment of acute infection. Nature 407 (6803), 523 (2000).
37. Trkola, A. et al., Delay of HIV-1 rebound after cessation of antiretroviral therapy through passive transfer of human neutralizing antibodies. Nat Med 11 (6), 615 (2005).
38. Fomsgaard, A. et al., Full-length characterization of A1/D intersubtype recombinant genomes from a therapy-induced HIV type 1 controller during acute infection and his noncontrolling partner. AIDS Res Hum Retroviruses 24 (3), 463 (2008).
39. Worm, S. W. et al., Risk of myocardial infarction in patients with HIV infection exposed to specific individual antiretroviral drugs from the 3 major drug classes: the data collection on adverse events of anti-HIV drugs (D:A:D) study. J Infect Dis 201 (3), 318 (2010).
40. Elzi, L. et al., Treatment modification in human immunodeficiency virus-infected individuals starting combination antiretroviral therapy between 2005 and 2008. Arch Intern Med 170 (1), 57 (2010).
41. Nguyen, A. et al., Lipodystrophy and weight changes: data from the Swiss HIV Cohort Study, 2000-2006. HIV Med 9 (3), 142 (2008).
42. Lowy, A. et al., Costs of treatment of Swiss patients with HIV on antiretroviral therapy in hospital-based and general practice-based care: a prospective cohort study. AIDS Care 17 (6), 698 (2005).
43. Farnham, P. G., Do Reduced Inpatient Costs Associated with Highly Active Antiretroviral Therapy (HAART) Balance the Overall Cost for HIV Treatment? Appl Health Econ Health Policy 8 (2), 75 (2010).
44. Ehrlich, P., Chemotherapeutics: Scientific Principles, Methods, and Results. Lancet 2, 445 (1913).
45. Cohn, M. L., Middlebrook, G., and Russell, W. F., Jr., Combined drug treatment of tuberculosis. I. Prevention of emergence of mutant populations of tubercle bacilli resistant to both streptomycin and isoniazid in vitro. J Clin Invest 38 (8), 1349 (1959).
46. Strain, M. C. et al., Effect of treatment, during primary infection, on establishment and clearance of cellular reservoirs of HIV-1. J Infect Dis 191 (9), 1410 (2005).
47. Altfeld, M. et al., Cellular immune responses and viral diversity in individuals treated during acute and early HIV-1 infection. J Exp Med 193 (2), 169 (2001).
48. Delwart, E. et al., Homogeneous quasispecies in 16 out of 17 individuals during very early HIV-1 primary infection. AIDS 16 (2), 189 (2002).
49. Sagar, M. et al., Infection with multiple human immunodeficiency virus type 1 variants is associated with faster disease progression. J Virol 77 (23), 12921 (2003).
50. Joos, B. et al., Low human immunodeficiency virus envelope diversity correlates with low in vitro replication capacity and predicts spontaneous control of plasma viremia after treatment interruptions. J Virol 79 (14), 9026 (2005).
51. Jost, S. et al., A patient with HIV-1 superinfection. N Engl J Med 347 (10), 731 (2002).
52. Yang, O. O. et al., Human immunodeficiency virus type 1 clade B superinfection: evidence for differential immune containment of distinct clade B strains. J Virol 79 (2), 860 (2005).
53. Streeck, H. et al., Immune-driven recombination and loss of control after HIV superinfection. J Exp Med 205 (8), 1789 (2008).

54. Clerc, O., Colombo, S., Yerly, S., Telenti, A., and Cavassini, M., HIV-1 elite controllers: Beware of super-infections. J Clin Virol 47, 376 (2010).
55. Altfeld, M. et al., HIV-1 superinfection despite broad CD8+ T-cell responses containing replication of the primary virus. Nature 420 (6914), 434 (2002).
56. Smith, D. M., Richman, D. D., and Little, S. J., HIV superinfection. J Infect Dis 192 (3), 438 (2005).
57. Fischer, M. et al., Cellular viral rebound after cessation of potent antiretroviral therapy predicted by levels of multiply spliced HIV-1 RNA encoding nef. J Infect Dis 190 (11), 1979 (2004).
58. Althaus, C. F. et al., Rational design of HIV-1 fluorescent hydrolysis probes considering phylogenetic variation and probe performance. J Virol Methods 165 (2), 151 (2010).
59. Lassen, K. G., Ramyar, K. X., Bailey, J. R., Zhou, Y., and Siliciano, R. F., Nuclear retention of multiply spliced HIV-1 RNA in resting CD4+ T cells. PLoS Pathog 2 (7), e68 (2006).
60. Rieder, P. et al., HIV-1 transmission after cessation of early antiretroviral therapy among men having sex with men. AIDS (2010).
61. Furtado, M. R. et al., Persistence of HIV-1 transcription in peripheral-blood mononuclear cells in patients receiving potent antiretroviral therapy. N Engl J Med. 340, 1614 (1999).
62. Metzner, K. J. et al., Efficient suppression of minority drug-resistant HIV type 1 (HIV-1) variants present at primary HIV-1 infection by ritonavir-boosted protease inhibitor-containing antiretroviral therapy. J Infect Dis 201 (7), 1063 (2010).
63. Fischer, M. et al., Residual HIV-RNA levels persist for up to 2.5 years in PBMC of patients on potent antiretroviral therapy. AIDS Res Hum Retroviruses 16 (12), 1135 (2000).
64. Otero, M. et al., Peripheral blood Dendritic cells are not a major reservoir for HIV type 1 in infected individuals on virally suppressive HAART. AIDS Res Hum Retroviruses 19 (12), 1097 (2003).
65. Pasternak, A. O. et al., Cellular levels of HIV unspliced RNA from patients on combination antiretroviral therapy with undetectable plasma viremia predict the therapy outcome. PLoS One 4 (12), e8490 (2009).
66. Guadalupe, M. et al., Viral suppression and immune restoration in the gastrointestinal mucosa of human immunodeficiency virus type 1-infected patients initiating therapy during primary or chronic infection. J Virol 80 (16), 8236 (2006).
67. George, M. D., Reay, E., Sankaran, S., and Dandekar, S., Early antiretroviral therapy for simian immunodeficiency virus infection leads to mucosal CD4+ T-cell restoration and enhanced gene expression regulating mucosal repair and regeneration. J Virol 79 (5), 2709 (2005).
68. Spina, C. A., Kwoh, T. J., Chowers, M. Y., Guatelli, J. C., and Richman, D. D., The importance of nef in the induction of human immunodeficiency virus type 1 replication from primary quiescent CD4 lymphocytes. J Exp Med 179 (1), 115 (1994).
69. Feinberg, M. B., Changing the natural history of HIV disease. Lancet 348 (9022), 239 (1996).
70. von Wyl, V. et al., Emergence of HIV-1 drug resistance in previously untreated patients initiating combination antiretroviral treatment: a comparison of different regimen types. Arch Intern Med 167 (16), 1782 (2007).

71. Huber, M. et al., Complement lysis activity in autologous plasma is associated with lower viral loads during the acute phase of HIV-1 infection. PLoS Med 3 (11), e441 (2006).
72. Christopherson, C. et al., PCR-Based assay to quantify human immunodeficiency virus type 1 DNA in peripheral blood mononuclear cells. J Clin Microbiol 38 (2), 630 (2000).
73. Roscic-Mrkic, B. et al., RANTES (CCL5) utilizes the proteoglycan CD44 as an auxiliary receptor to mediate cellular activation signals and HIV-1 enhancement. Blood 102 (4), 1169 (2003).

Conclusions and Outlook

Combination antiretroviral therapy has tremendously reduced morbidity and mortality of HIV-1 infected patients[4]. Nevertheless, HIV-1 infection can not be cured. The main obstacle is the presence of latently infected cells which contain proviral DNA that is transcriptionally not active, however, may be activated and produce virions upon certain stimuli[37,40,45]. The three projects described in this thesis approach different aspects of HIV-1 latency.

In the first project, prerequisites for the optimal design of qPCR for phylogenetically diverse organisms, such as HIV-1[108], were defined. Highly sensitive qPCR is particularly needed to quantify RNA transcripts in latently HIV-1 infected cells from patients, successfully treated with cART. Primers can be relatively forgiving in respect to target variation[109], but also their optimal design and target complementarity should be considered for highly sensitive qPCR assays. However, we were focusing on the fluorescent hydrolysis probes (FH-probes), as they are very sensitive to mismatches in their target recognition site[110], which could lead to an underestimation of RNA copy numbers in the case of highly variable HIV-1. On the other hand, they are the most sensitive reagents to quantify specifically genome copy numbers by qPCR[111]. Thus, more than 50 FH-probes have been investigated in terms of their performance in qPCR assays. Our data allowed to determine positive predictors for optimal functionality, for instance, a purine at the second position of the 5' end, a high melting temperature of the 5' end, and the presence of G/C rich secondary structures predict a efficient FH-probe performance in qPCR assays[112]. Additionally, application of an algorithm to scan sequence databases for FH-probes with optimal phylogenetic conservation enabled the identification of functional FH-probes in various regions of the HIV-1 genome approaching coverage of the global HIV-1 pandemic. The established phylogenetic and biochemical principles for FH-probe and qPCR design may also be extended to other phylogenetically diverse biological systems.

A first application of those highly sensitive qPCR assays is shown in the third project. HIV-1 transcription was monitored in PBMC of acutely HIV-1 infected patients to determine the effects of early treatment. Especially for measuring the different HIV-1 transcripts in patients under successful treatment single copy sensitive PCRs are needed. Interestingly, cART initiated during acute HIV-1 infection significantly

depleted the number of transcriptionally active proviruses by an order of magnitude when compared to levels detected in patients starting cART during chronic HIV-1 infection[113].

In summary, highly sensitive qPCRs are essential when very low levels of the different HIV-1 transcripts are expected. Also further studies, analyzing possible latency mechanisms and treatment options to cure HIV-1, will depend on highly sensitive PCR assays.

In the second project, a novel method was established to enrich low abundant, small noncoding RNAs (sncRNAs). HIV-1 sncRNAs are generated in very low frequencies in HIV-1 infected cells, if at all (0-<0.5%)[80,98,107]. So far, only a few HIV-1 derived sncRNAs have been described[82,84,85,101,102]. We hypothesize that HIV-1 latency might be induced by RNA interference, a post-transcriptional silencing pathway ubiquitously present in eukaryotic cells of many organisms and mediated by sncRNAs[114]. Thus, the first step to prove this hypothesis was the development of an enrichment method to detect HIV-1 derived sncRNAs. More than 100-fold enrichment could be achieved using our hybridization capture based enrichment protocol, which includes two rounds of selection where HIV-1 derived sncRNAs bind specifically to long HIV-1 ssDNA hybridization probes covering the whole virus genome. This method was applied to several HIV-1 infected primary macrophages and $CD4^+$ T-lymphocytes showing its reproducibility, sensitivity, and specificity. SncRNA libraries obtained with our method contained more than 70% of HIV-1 derived sncRNAs. While high-throughput sequencing techniques certainly have the capacity to overcome the limitations in identifying low abundant sncRNAs, it must still be considered that more than 99% of sequenced sncRNAs retrieved by random sequencing will not be of interest and very low abundant sncRNAs might still be missed. Our approach allows sequence specific selection with a high sensitivity, i.e., all captured sncRNAs are of potential interest even if they are not derived from HIV-1. Certainly, also a combined approach, hybridization capture followed by high-throughput sequencing instead of molecular cloning and Sanger sequencing, should be conceivable and would combine advantages of both methods.

With this highly optimized method, 216 unique, HIV-1 derived clones were identified from a total of 9 sncRNA libraries. Those were distributed throughout the viral

genome; however, they tend to cluster in certain regions of the viral genome where these >200 sncRNAs formed 67 contigs. An obvious question that arises is whether the HIV-1 sncRNAs detected here are not just mere degradation products. There are several points arguing against it. Firstly, the HIV-1 sncRNAs are not randomly distributed as it would be expected in case of degraded products, but are rather localized in certain hot-spots at distinct sites. One could still argue that certain degraded RNA fragments are protected from further degradation due to their secondary structure. However, no evidence for preferential clustering in highly structured regions of the HIV-1 genome[10] were observed (except HIV-1 sncRNAs within the TAR region). Secondly, the enzymes used for C-tailing and adaptor ligation require specific ends, namely a phosphate at the 5' end and a hydroxyl at the 3' end which are not produced by the common RNA degradation process via hydrolysis[115]. Thirdly, HIV-1 sncRNAs were identified that correspond to the regions where HIV-1 sncRNAs (hiv-1-miR-TAR-3p, hiv-1-miR-H1, and hiv1-miR-N367) have been described previously[84,85,102].

At last, further characterization of these HIV-1 derived sncRNAs was started. The presence of 4 sncRNAs contigs from different regions of the viral genome was confirmed in cells infected with primary virus isolates derived from 5 different HIV-1 patients, which excluded the possibility that these sncRNAs are specific for the virus isolate HIV-1_{JRFL} used for all former experiments. Additionally, functional assays were performed by transfecting primary macrophages with two HIV-1 sncRNAs contigs. This identified one contig, consisting of sense and antisense sncRNAs, which efficiently inhibited HIV-1 replication.

The next steps will be to further characterize the functions of the identified HIV-1 sncRNAs as well as their biogenesis pathways. Most of the clones are in sense direction and therefore not able to target the plus-sense, single-stranded RNA genome of HIV-1 at the accordant position. Nevertheless, they might target other positions in the HIV-1 genome or human cellular RNA transcripts. Interestingly, approximately 10% of the HIV-1 sncRNAs are in antisense direction, thus, they could hybridize to the HIV-1 RNA genome and inhibit virus replication via the RNA interference pathway. Besides investigating their potential as HIV-1 inhibitors, it needs to be explored how antisense sncRNAs are generated. It is generally assumed that HIV-1 antisense transcripts are not generated[14], however, there are few reports showing that antisense transcripts are produced in low levels[116-118]. Those transcripts

might either be initiated by the HIV-1 promoter located in the 3' LTR[119] or by a downstream cellular promoter[120]. Nevertheless, it is still unknown whether they are just random by-products, or if they have specific functions[121]. One possible function could be that they are processed to regulatory small RNAs, and lead to a downregulation of HIV-1 transcription, which could ultimately lead to HIV-1 latency. Another hypothesis is that they might upregulate HIV-1 replication, especially the sncRNAs targeting the promoter region[122,123], and therefore rather present a way for the virus to control its reactivation after latency . In case those sncRNAs would have opposite effects, this would open a variety of options for developing novel treatment strategies.

To experimentally further address the potential functions of the HIV-1 sncRNAs, HIV-1 sncRNA transfection/HIV-1 infection experiment should be performed to asses their impact on HIV-1 replication. Furthermore, their interaction with, for instance, Ago or PIWI could be investigated by immunoprecipitation. For HIV-1 sncRNAs having an impact on HIV-1 replication and interacting with RNAi pathway proteins, an *in silico* target screening could be performed on both the viral and the cellular mRNA transcripts.

In conclusion, our method presents a validated and very efficient approach to enrich and identify novel, low abundant HIV-1 sncRNAs. HIV-1 derived sncRNAs are a very exciting field, which could have an enormous impact on novel therapy designs. There are currently numerous siRNAs tested for treatment of, for instant, cancer[124,125] and infectious diseases[126] (clinicaltrials.gov).

In summary, our selection method has an enormous capacity to enrich low abundant sncRNAs in a sequence specific manner, which highly recommends its usage also for the detection of sncRNAs of other origins. The method can be applied for any type of investigations where sncRNAs from known origins or hybridizing to a known target sequence are sought for.

Abbreviations

AATF	apoptosis antagonizing transcription factor
Ago	Argonaut
AIDS	acquired immunodeficiency syndrome
APOBEG3G	apoliprotein B mRNA editing enzyme, catalytic polypeptide-like 3G
bp	base pair
CA	capsid
cART	combination antiretroviral therapy
cDNA	complementary DNA
Ct	treshhold cycle
Delta-Tm	Tm difference of FH-probe and the primer binding the same strand
DIS	dimerization initiation site
DLNA	DNA locked nucleic acid
DNA	deoxyribonucleic acid
dsRNA	double-stranded RNA
EBV	Epstein-Barr virus
Env	envelope
FH-probe	fluorescent hydrolysis probe
Gag	group specific antigen
GALT	gut associated lymphoid tissue
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
HAT	Histone acetyltransferase
HDAC1	Histone deacetylase 1
HIV	Human immunodeficiency virus
IN	integrase
KSHV	Kaposi’s sarcoma-associated herpesvirus
LTR	Long Terminal Repeat
MA	matrix
miRNA	microRNA
mRNA	messenger RNA
MsRNA	multiply spliced RNA
NC	nucleocapsid
ncRNA	noncoding RNA

Nef	negative replication factor
NF-κB	nuclear factor kappa b
nt	nucleotide
PBMC	peripheral blood mononuclear cell
PBS	primer binding site
PCAF	p300/CBP-associated factor
PCR	polymerase chain reaction
PHA	phytohaemagglutinin
piRNA	Piwi-interacting RNA
Pol	polymerase
PR	protease
Pre-miRNA	precursor microRNA
P-TEFb	Positive transcription elongation factor b
qPCR	quantitative polymerase chain reaction
R	repeat region in the LTR
Rev	Regulator of virion expression
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
RRE	Rev responsive element
rRNA	ribosomal RNA
RT	reverse transcriptase
RT-PCR	reverse transcription-PCR
sA7	splice acceptor 7
SHCS	Swiss HIV Cohort Study
siRNA	small interfering RNA
sncRNA	small noncoding RNA
SNR	signal to noise ratio
ssRNA	single stranded RNA
SU	surface glycoprotein
TAR	trans-activating response region
Tat	transactivator of transcription
Tm	melting temperature
TM	transmembrane glycoprotein

TRBP	TAR RNA binding protein
tRNA	transfer RNA
TSA	trichostation A
UsRNA	unspliced RNA
vDNA	viral DNA
vmiRNA	viral micro RNA
vsiRNA	viral small interfering RNA
VPA	valproic acid
Vpr	Viral protein R
Vpu	Viral protein
vRex	PBMC-associated extracellular virion RNA
ZPHI	Zurich Primary HIV Study

References

1. Hymes, K. B. et al., Kaposi's sarcoma in homosexual men-a report of eight cases. Lancet 2, 598 (1981).
2. Gottlieb, M. S. et al., Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N Engl J Med 305, 1425 (1981).
3. Barre-Sinoussi, F. et al., Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220, 868 (1983).
4. Sterne, J. A. et al., Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. Lancet 366, 378 (2005).
5. Clavel, F. et al., Isolation of a new human retrovirus from West African patients with AIDS. Science 233, 343 (1986).
6. Marlink, R. et al., Reduced rate of disease development after HIV-2 infection as compared to HIV-1. Science 265, 1587 (1994).
7. Kurth, R. and Bannert, N., Retroviruses: Molecular Biology, Genomics and Pathogenesis (Caister Academic Press, 2010).
8. Kao, S. Y., Calman, A. F., Luciw, P. A., and Peterlin, B. M., Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature 330, 489 (1987).
9. Laspia, M. F., Rice, A. P., and Mathews, M. B., HIV-1 Tat protein increases transcriptional initiation and stabilizes elongation. Cell 59, 283 (1989).
10. Watts, J. M. et al., Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460, 711 (2009).
11. Ratner, L. et al., Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature 313, 277 (1985).
12. Jiang, M. et al., Identification of tRNAs incorporated into wild-type and mutant human immunodeficiency virus type 1. J Virol 67, 3246 (1993).
13. McLaren, M., Marsh, K., and Cochrane, A., Modulating HIV-1 RNA processing and utilization. Front Biosci 13, 5693 (2008).
14. Frankel, A. D. and Young, J. A., HIV-1: fifteen proteins and an RNA. Annu Rev Biochem 67, 1 (1998).
15. Hallenberger, S. et al., Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. Nature 360, 358 (1992).
16. Daly, T. J., Cook, K. S., Gray, G. S., Maione, T. E., and Rusche, J. R., Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro. Nature 342, 816 (1989).
17. Zapp, M. L. and Green, M. R., Sequence-specific RNA binding by the HIV-1 Rev protein. Nature 342, 714 (1989).
18. Malim, M. H., Hauber, J., Le, S. Y., Maizel, J. V., and Cullen, B. R., The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature 338, 254 (1989).
19. Foster, J. L. and Garcia, J. V., HIV-1 Nef: at the crossroads. Retrovirology 5, 84 (2008).
20. Goila-Gaur, R. and Strebel, K., HIV-1 Vif, APOBEC, and intrinsic immunity. Retrovirology 5, 51 (2008).
21. Ganser-Pornillos, B. K., Yeager, M., and Sundquist, W. I., The structural biology of HIV assembly. Curr Opin Struct Biol 18, 203 (2008).

22. Moore, J. P., Trkola, A., and Dragic, T., Co-receptors for HIV-1 entry. Curr Opin Immunol 9, 551 (1997).
23. Suzuki, Y. and Craigie, R., The road to chromatin - nuclear entry of retroviruses. Nat Rev Microbiol 5, 187 (2007).
24. Tang, H., Kuhen, K. L., and Wong-Staal, F., Lentivirus replication and regulation. Annu Rev Genet 33, 133 (1999).
25. Pollard, V. W. and Malim, M. H., The HIV-1 Rev protein. Annu Rev Microbiol 52, 491 (1998).
26. Dingwall, C. et al., Human immunodeficiency virus 1 tat protein binds trans-activation-responsive region (TAR) RNA in vitro. Proc Natl Acad Sci U S A 86, 6925 (1989).
27. Adamson, C. S. and Freed, E. O., Human immunodeficiency virus type 1 assembly, release, and maturation. Adv Pharmacol 55, 347 (2007).
28. Coiras, M., Lopez-Huertas, M. R., Perez-Olmeda, M., and Alcami, J., Understanding HIV-1 latency provides clues for the eradication of long-term reservoirs. Nat Rev Microbiol 7, 798 (2009).
29. Pope, M. and Haase, A. T., Transmission, acute HIV-1 infection and the quest for strategies to prevent infection. Nat Med 9, 847 (2003).
30. Hoffmann, C., Rockstroh, J., and Kamps, B., HIV Medicine 2007, 15th edition ed. (Flying Publisher, 2007).
31. Aceto, L. et al., [Primary HIV-1 infection in Zurich: 2002-2004]. Schweiz Rundsch Med Prax 94, 1199 (2005).
32. Frost, S. D., Trkola, A., Gunthard, H. F., and Richman, D. D., Antibody responses in primary HIV-1 infection. Curr Opin HIV AIDS 3, 45 (2008).
33. Fischer, M. et al., Residual HIV-RNA levels persist for up to 2.5 years in PBMC of patients on potent antiretroviral therapy. AIDS Res.Hum.Retroviruses 16, 1135 (2000).
34. Furtado, M. R. et al., Persistence of HIV-1 transcription in peripheral-blood mononuclear cells in patients receiving potent antiretroviral therapy. N.Engl.J Med. 340, 1614 (1999).
35. Finzi, D. et al., Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. Nat Med 5, 512 (1999).
36. Blankson, J. N., Persaud, D., and Siliciano, R. F., The challenge of viral reservoirs in HIV-1 infection. Annu.Rev.Med. 53, 557 (2002).
37. Joos, B. et al., HIV rebounds from latently infected cells, rather than from continuing low-level replication. Proc Natl Acad Sci U S A 105, 16725 (2008).
38. Mackowiak, P. A., Microbial latency. Rev Infect Dis 6, 649 (1984).
39. Weinberger, L. S. and Shenk, T., An HIV feedback resistor: auto-regulatory circuit deactivator and noise buffer. PLoS Biol 5, e9 (2006).
40. Lassen, K., Han, Y., Zhou, Y., Siliciano, J., and Siliciano, R. F., The multifactorial nature of HIV-1 latency. Trends Mol Med 10, 525 (2004).
41. Marzio, G., Tyagi, M., Gutierrez, M. I., and Giacca, M., HIV-1 tat transactivator recruits p300 and CREB-binding protein histone acetyltransferases to the viral promoter. Proc Natl Acad Sci U S A 95, 13519 (1998).
42. He, G. and Margolis, D. M., Counterregulation of chromatin deacetylation and histone deacetylase occupancy at the integrated promoter of human immunodeficiency virus type 1 (HIV-1) by the HIV-1 repressor YY1 and HIV-1 activator Tat. Mol Cell Biol 22, 2965 (2002).
43. Van Duyne, R. et al., Lysine methylation of HIV-1 Tat regulates transcriptional activity of the viral LTR. Retrovirology 5, 40 (2008).

44. Lusic, M., Marcello, A., Cereseto, A., and Giacca, M., Regulation of HIV-1 gene expression by histone acetylation and factor recruitment at the LTR promoter. EMBO J 22, 6550 (2003).
45. Ylisastigui, L., Archin, N. M., Lehrman, G., Bosch, R. J., and Margolis, D. M., Coaxing HIV-1 from resting CD4 T cells: histone deacetylase inhibition allows latent viral expression. Aids 18, 1101 (2004).
46. Ott, M. et al., Acetylation of the HIV-1 Tat protein by p300 is important for its transcriptional activity. Curr Biol 9, 1489 (1999).
47. Kiernan, R. E. et al., HIV-1 tat transcriptional activity is regulated by acetylation. EMBO J 18, 6106 (1999).
48. Nabel, G. and Baltimore, D., An inducible transcription factor activates expression of human immunodeficiency virus in T cells. Nature 326, 711 (1987).
49. Corthesy, B. and Kao, P. N., Purification by DNA affinity chromatography of two polypeptides that contact the NF-AT DNA binding site in the interleukin 2 promoter. J Biol Chem 269, 20682 (1994).
50. Kinter, A. L., Poli, G., Maury, W., Folks, T. M., and Fauci, A. S., Direct and cytokine-mediated activation of protein kinase C induces human immunodeficiency virus expression in chronically infected promonocytic cells. J Virol 64, 4306 (1990).
51. Jakobovits, A., Rosenthal, A., and Capon, D. J., Trans-activation of HIV-1 LTR-directed gene expression by tat requires protein kinase C. EMBO J 9, 1165 (1990).
52. Weil, R. and Israel, A., T-cell-receptor- and B-cell-receptor-mediated activation of NF-kappaB in lymphocytes. Curr Opin Immunol 16, 374 (2004).
53. Coiras, M., Lopez-Huertas, M. R., Rullas, J., Mittelbrunn, M., and Alcami, J., Basal shuttle of NF-kappaB/I kappaB alpha in resting T lymphocytes regulates HIV-1 LTR dependent expression. Retrovirology 4, 56 (2007).
54. Kinoshita, S. et al., The T cell activation factor NF-ATc positively regulates HIV-1 replication and gene expression in T cells. Immunity 6, 235 (1997).
55. Garber, M. E., Wei, P., and Jones, K. A., HIV-1 Tat interacts with cyclin T1 to direct the P-TEFb CTD kinase complex to TAR RNA. Cold Spring Harb Symp Quant Biol 63, 371 (1998).
56. Zhou, M. et al., Tat modifies the activity of CDK9 to phosphorylate serine 5 of the RNA polymerase II carboxyl-terminal domain during human immunodeficiency virus type 1 transcription. Mol Cell Biol 20, 5077 (2000).
57. Kessler, M. and Mathews, M. B., Premature termination and processing of human immunodeficiency virus type 1-promoted transcripts. J Virol 66, 4488 (1992).
58. Toohey, M. G. and Jones, K. A., In vitro formation of short RNA polymerase II transcripts that terminate within the HIV-1 and HIV-2 promoter-proximal downstream regions. Genes Dev 3, 265 (1989).
59. Richman, D. D. et al., The challenge of finding a cure for HIV infection. Science 323, 1304 (2009).
60. Ecker, J. R. and Davis, R. W., Inhibition of gene expression in plant cells by expression of antisense RNA. Proc Natl Acad Sci U S A 83, 5372 (1986).
61. Fire, A. et al., Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391, 806 (1998).
62. Ghildiyal, M. and Zamore, P. D., Small silencing RNAs: an expanding universe. Nat Rev Genet 10, 94 (2009).

63. Lee, R. C., Feinbaum, R. L., and Ambros, V., The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75, 843 (1993).
64. Ambros, V., The functions of animal microRNAs. Nature 431, 350 (2004).
65. Aravin, A. A. et al., The small RNA profile during Drosophila melanogaster development. Dev Cell 5, 337 (2003).
66. Olivieri, D., Sykora, M. M., Sachidanandam, R., Mechtler, K., and Brennecke, J., An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. EMBO J 29, 3301 (2010).
67. Kim, V. N., Han, J., and Siomi, M. C., Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol 10, 126 (2009).
68. Gregory, R. I. et al., The Microprocessor complex mediates the genesis of microRNAs. Nature 432, 235 (2004).
69. Han, J. et al., The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev 18, 3016 (2004).
70. Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E., and Kutay, U., Nuclear export of microRNA precursors. Science 303, 95 (2004).
71. Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J., Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409, 363 (2001).
72. Schwarz, D. S. et al., Asymmetry in the assembly of the RNAi enzyme complex. Cell 115, 199 (2003).
73. Rana, T. M., Illuminating the silence: understanding the structure and function of small RNAs. Nat Rev Mol Cell Biol 8, 23 (2007).
74. Suzuki, K. and Kelleher, A. D., Transcriptional regulation by promoter targeted RNAs. Curr Top Med Chem 9, 1079 (2009).
75. Grivna, S. T., Beyret, E., Wang, Z., and Lin, H., A novel class of small RNAs in mouse spermatogenic cells. Genes Dev 20, 1709 (2006).
76. Grivna, S. T., Pyhtila, B., and Lin, H., MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. Proc Natl Acad Sci U S A 103, 13415 (2006).
77. Grosshans, H. and Filipowicz, W., Molecular biology: the expanding world of small RNAs. Nature 451, 414 (2008).
78. Pfeffer, S. et al., Identification of virus-encoded microRNAs. Science 304, 734 (2004).
79. Grey, F. et al., Identification and characterization of human cytomegalovirus-encoded microRNAs. J Virol 79, 12095 (2005).
80. Pfeffer, S. et al., Identification of microRNAs of the herpesvirus family. Nat Methods 2, 269 (2005).
81. Sullivan, C. S., Grundhoff, A. T., Tevethia, S., Pipas, J. M., and Ganem, D., SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. Nature 435, 682 (2005).
82. Bennasser, Y., Le, S. Y., Benkirane, M., and Jeang, K. T., Evidence that HIV-1 encodes an siRNA and a suppressor of RNA silencing. Immunity 22, 607 (2005).
83. Omoto, S. and Fujii, Y. R., Regulation of human immunodeficiency virus 1 transcription by nef microRNA. J Gen Virol 86, 751 (2005).
84. Omoto, S. et al., HIV-1 nef suppression by virally encoded microRNA. Retrovirology 1, 44 (2004).

85. Ouellet, D. L. et al., Identification of functional microRNAs released through asymmetrical processing of HIV-1 TAR element. Nucleic Acids Res 36, 2353 (2008).
86. Cai, X. et al., Kaposi's sarcoma-associated herpesvirus expresses an array of viral microRNAs in latently infected cells. Proc Natl Acad Sci U S A 102, 5570 (2005).
87. Lo, A. K. et al., Modulation of LMP1 protein expression by EBV-encoded microRNAs. Proc Natl Acad Sci U S A 104, 16164 (2007).
88. Xing, L. and Kieff, E., Epstein-Barr virus BHRF1 micro- and stable RNAs during latency III and after induction of replication. J Virol 81, 9967 (2007).
89. Narayanan, A., Kehn-Hall, K., Bailey, C., and Kashanchi, F., Analysis of the roles of HIV-derived microRNAs. Expert Opin Biol Ther 11, 17 (2011).
90. Yeung, M. L. et al., Changes in microRNA expression profiles in HIV-1-transfected human cells. Retrovirology 2, 81 (2005).
91. Triboulet, R. et al., Suppression of microRNA-silencing pathway by HIV-1 during virus replication. Science 315, 1579 (2007).
92. Sung, T. L. and Rice, A. P., miR-198 inhibits HIV-1 gene expression and replication in monocytes and its mechanism of action appears to involve repression of cyclin T1. PLoS Pathog 5, e1000263 (2009).
93. Ahluwalia, J. K. et al., Human cellular microRNA hsa-miR-29a interferes with viral nef protein expression and HIV-1 replication. Retrovirology 5, 117 (2008).
94. Huang, J. et al., Cellular microRNAs contribute to HIV-1 latency in resting primary CD4+ T lymphocytes. Nat Med 13, 1241 (2007).
95. Wang, X. et al., Cellular microRNA expression correlates with susceptibility of monocytes/macrophages to HIV-1 infection. Blood 113, 671 (2009).
96. Li, W. X. et al., Interferon antagonist proteins of influenza and vaccinia viruses are suppressors of RNA silencing. Proc Natl Acad Sci U S A 101, 1350 (2004).
97. Lecellier, C. H. et al., A cellular microRNA mediates antiviral defense in human cells. Science 308, 557 (2005).
98. Lin, J. and Cullen, B. R., Analysis of the interaction of primate retroviruses with the human RNA interference machinery. J Virol 81, 12218 (2007).
99. Sanghvi, V. R. and Steel, L. F., A Re-Examination of Global Suppression of RNA Interference by HIV-1. PLoS One 6, e17246 (2011).
100. Bennasser, Y., Le, S. Y., Yeung, M. L., and Jeang, K. T., HIV-1 encoded candidate micro-RNAs and their cellular targets. Retrovirology 1, 43 (2004).
101. Klase, Z. et al., HIV-1 TAR element is processed by Dicer to yield a viral micro-RNA involved in chromatin remodeling of the viral LTR. BMC Mol Biol 8, 63 (2007).
102. Kaul, D., Khanna, A., and Suman, Evidence and nature of a novel miRNA encoded by HIV-1. Proc Indian Natl Sci Acad 72, 91 (2006).
103. Klase, Z. et al., HIV-1 TAR miRNA protects against apoptosis by altering cellular gene expression. Retrovirology 6, 18 (2009).
104. Purzycka, K. J. and Adamiak, R. W., The HIV-2 TAR RNA domain as a potential source of viral-encoded miRNA. A reconnaissance study. Nucleic Acids Symp Ser (Oxf), 511 (2008).
105. Kaul, D., Ahlawat, A., and Gupta, S. D., HIV-1 genome-encoded hiv1-mir-H1 impairs cellular responses to infection. Mol Cell Biochem 323, 143 (2009).
106. Lamers, S. L., Fogel, G. B., and McGrath, M. S., HIV-miR-H1 evolvability during HIV pathogenesis. Biosystems 101, 88 (2010).

107. Yeung, M. L. et al., Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid. Nucleic Acids Res 37, 6575 (2009).
108. Buonaguro, L., Tornesello, M. L., and Buonaguro, F. M., Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. J Virol 81, 10209 (2007).
109. Christopherson, C., Sninsky, J., and Kwok, S., The effects of internal primer-template mismatches on RT-PCR: HIV-1 model studies. Nucleic Acids Res 25, 654 (1997).
110. Damond, F. et al., Human immunodeficiency virus type 1 (HIV-1) plasma load discrepancies between the Roche COBAS AMPLICOR HIV-1 MONITOR Version 1.5 and the Roche COBAS AmpliPrep/COBAS TaqMan HIV-1 assays. J Clin Microbiol 45, 3436 (2007).
111. Wang, L., Blasic, J. R., Jr., Holden, M. J., and Pires, R., Sensitivity comparison of real-time PCR probe designs on a model DNA plasmid. Anal Biochem 344, 257 (2005).
112. Althaus, C. F. et al., Rational design of HIV-1 fluorescent hydrolysis probes considering phylogenetic variation and probe performance. J Virol Methods 165, 151 (2010).
113. Schmid, A. et al., Profound depletion of HIV-1 transcription in patients initiating antiretroviral therapy during acute infection. PLoS One 5, e13310 (2010).
114. Farazi, T. A., Juranek, S. A., and Tuschl, T., The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. Development 135, 1201 (2008).
115. Landgraf, P. et al., A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129, 1401 (2007).
116. Michael, N. L. et al., Negative-strand RNA transcripts are produced in human immunodeficiency virus type 1-infected cells and patients by a novel promoter downregulated by Tat. J Virol 68, 979 (1994).
117. Landry, S. et al., Detection, characterization and regulation of antisense transcripts in HIV-1. Retrovirology 4, 71 (2007).
118. Vanhee-Brossollet, C. et al., A natural antisense RNA derived from the HIV-1 env gene encodes a protein which is recognized by circulating antibodies of HIV+ individuals. Virology 206, 196 (1995).
119. Ludwig, L. B. et al., Human Immunodeficiency Virus-Type 1 LTR DNA contains an intrinsic gene producing antisense RNA and protein products. Retrovirology 3, 80 (2006).
120. Lenasi, T., Contreras, X., and Peterlin, B. M., Transcriptional interference antagonizes proviral gene expression to promote HIV latency. Cell Host Microbe 4, 123 (2008).
121. Faghihi, M. A. and Wahlestedt, C., Regulatory roles of natural antisense transcripts. Nat Rev Mol Cell Biol 10, 637 (2009).
122. Janowski, B. A. et al., Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. Nat Chem Biol 3, 166 (2007).
123. Li, L. C. et al., Small dsRNAs induce transcriptional activation in human cells. Proc Natl Acad Sci U S A 103, 17337 (2006).
124. Petrocca, F. and Lieberman, J., Promise and challenge of RNA interference-based therapy for cancer. J Clin Oncol 29, 747 (2011).
125. Davis, M. E. et al., Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. Nature 464, 1067 (2010).

126. Lanford, R. E. et al., Therapeutic silencing of microRNA-122 in primates with chronic hepatitis C virus infection. Science 327, 198 (2010).

Acknowledgements

First, I would like to express my deepest gratitude to Marek Fischer. He was the best supervisor I could have wished for, and always extremely supportive of his students. I will always remember his patience and kindness.

I am also very grateful to Karin Metzner. She decided to “adopt” Valentina and me, even though this meant a lot of additional work for her. Nevertheless, she always found time for everyone. Thank you very much for everything.

Many thanks go to Huldrych Günthard, who always supported our project with much effort and many valuable inputs.

I would also like to thank the members of my thesis committee, Alexandra Trkola and John Robinson, for their constant support.

Endless thanks go to the other members of the “Fischer group”: Barbara, for her enormous support and help in the lab. Valentina, for being the only person who enjoyed talking about small RNAs and for her huge enthusiasm for our shared project, and to Beda, an inofficial member of our group, for always having time to assist us with sequence analysis and computer problems.

I would also like to thank all the present and former lab members for creating a wonderful familiar atmosphere. In particular Philip, who was a very reliable coffee-partner and became my best friend during our shared PhD years. Francesca, Christine, and Alexandra, for being not only colleagues but friends. Adrian, Sara, Lisa, Wan-Lin, Yannick, Li Duo, Marc and Ursi, with whom I shared many enjoyable moments.

Very special thanks go to my parents and my sister for always supporting and believing in me.

Curriculum Vitae

Personal Data

Name	CLAUDIA ALTHAUS
Date of birth	14. September 1982
Citizenship	Bretzwil BL

Education

03/2008-09/2011	PhD thesis at the University of Zurich, Thesis: *"Interplay between HIV-transcription and the pathways of intracellular innate defense."* Supervision: Prof. Dr. Marek Fischer, PD Dr. Karin Metzner, Prof. Dr. Alexandra Trkola, Prof. Dr. Huldrych Günthard and Prof. Dr. John Robinson
03/2007	Master of Science in Infection Biology and Epidemiology, University of Basel
10/2005 – 03/2007	Master thesis at the Swiss Tropical Institute, University of Basel Thesis: *"Establishment and validation of a new microfluorimetric method (PicoGreen) for Trypanosoma brucei gambiense"* Supervision: Prof. R. Brun
07/2005	Bachelor of Science in Biology, University of Basel
10/2002 – 07/2005	Studies of Animal and Plant Sciences, University of Basel
08/1998 –12/ 2001	Gymnasium (High School), Liestal BL, Type B

Publications

Althaus CF, Vongrad V, Niederöst B, Joos B, Di Giallonardo F, Rieder P, Trkola A Günthard HF, Metzner KJ & Fischer M. *Novel targeted enrichment strategy enables detection of low abundant small noncoding RNAs in HIV-1 infection.* Manuscript submitted

Schmid A, Gianella S, von Wyl V, Metzner KJ, Scherrer AU, Niederöst B, Althaus CF, Rieder P, Grube C, Joos B, Weber R, Fischer M, Günthard HF. *Profound depletion of HIV-1 transcription in patients initiating antiretroviral therapy during acute infection.* PLoS One. 2010 Oct 12;5(10)

Althaus CF, Gianella S, Rieder P, von Wyl V, Kouyos RD, Niederöst B, Schmid A, Metzner KJ, Joos B, Günthard HF, Fischer M. *Rational design of HIV-1 fluorescent hydrolysis probes considering phylogenetic variation and probe performance.* J Virol Methods. 2010 May;165(2):151-60